COVID-19 Public Tweets Sentiment Analysis using TF-IDF and Inductive Learning Models

RUBUL KUMAR BANIA

Dept. of Computer Application,
North-Eastern Hill University
Tura Campus, Meghalaya - 794002, India
rubul.bania@gmail.com
rubul.bania@nehu.ac.in

Abstract. The fact of appearing of the handheld devices offers forthright entry to the internet and social networking sites. Sentiment analysis and opinion mining is the study of sentiments or opinions shared by different users in social networking sites like, Twitter, Facebook, Reddit, Instagram etc., on diverse social phenomena. In this article, sentiment analysis of different tweets on the ongoing epidemic COVID-19, Corona virus disease is performed. COVID-19 is declared as epidemic by the World Health Organization (WHO) in the mid of March 2020. The statistical and machine learning based analyses are implemented on 40,000 tweets, which were collected in two different mutually exclusive time frames. Tweets are collected from Twitter site between 3/07/2020 to 11/07/2020 and 01/08/2020 to 06/08/2020, using Tweepy python library. Various Python based libraries are applied for data acquisition, data pre-processing and data analysis processes. As a data pre-processing phase initially sentences are cleaned. Then by calculating the polarity and subjectivity measures, tweets are categorized into three groups (viz., *negative*, *neutral* and *positive*). Thereafter, in the later phase by applying the Term frequency-inverse document frequency (TF-IDF) feature extraction scheme with the help of uni-gram, bi-gram, and tri-gram techniques different features are extracted to prepare the datasets to feed it into the prediction models. 70% of the datasets are used to train Gaussian Naïve Bayes (G-NB), Bernoulli's Naïve Bayes (B-NB), Random forest (RF), and Support vector machine (SVM) classifiers to generate different prediction models. Finally, 30% of the data is tested on those learning models. Experimental results suggest that RF and B-NB models performed better than the other two classifier models. The execution computational cost of SVM is very high.

Keywords: Sentiment analysis, Twitter, Covid-19, TF-IDF, Classification, Python.

(Received August 16th, 2020 / Accepted November 19th, 2020)

1 Introduction

Due to the technological revolutions and for the enrichment of World Wide Web, through out globe people are using social networking sites to put their opinion and sentiment about various issues or diverse social phenomena. Among the different social networking micro blogging sites like Facebook, Twitter, Reddit, Instagram, etc., Twitter is frequently used to tweets on various day to day worldwide activities [4]. Analysis of different forms of data from social networking sites are now a days very useful. Collecting images, videos and texts from such micro blogging sites are become a very good data repositories also. Sentiment analysis or opinion mining is used to detect the opinion or mood of different sentences which are in the form texts as *negative*, *neutral*, and *positive* [33, 7, 34, 36, 29, 13, 31, 28, 12, 27, 30, 20, 16]. These sentiments are very much important and can play a major role in predictions. Very recent coronavirus disease (COVID-19) pandemic [3] is taking a troll on the world's health care infrastructure as well as the social, economic, and psychological well-being of humanity. COVID-19 is declared as pandemic by WHO on 13 March, 2020 [3, 34, 10]. The first case of COVID-19 was reported in December 2019 from China. Different media and news articles assumed Wuhan city of China as the origin of this epidemic which become pandemic in the month of March 2020. The usual symptoms of the COVID-19 are very common like dry cough, fever, throat infection, difficulty in breathing etc. As per the studies there is no proper standard treatment and designated vaccines are identified. Though different trials on vaccine productions are going on in different countries. But the number of deaths and infected persons are increasing rapidly. As per WHO (https://covid19.who.int/) total 20,439,814 confirmed cases worldwide and 744,385 deaths, reported to WHO till 13 August, 2020. Hence this epidemic is becoming the biggest topic of concern for the entire world.

Now, through out globe, different individuals, different organizations, and government individuals are using social media to communicate with each other on a number of issues relating to the COVID-19 pandemic. But it is also sometimes very unfortunate that without knowing much about some sensitive issues, different topics being shared on social media platforms relating to COVID-19. Moreover, to understand the public sentiment in this epidemic situation is also desirable. That is why in this study COVID-19 epidemic is chosen for sentiments analysis from tweets. In a very recent study authors have highlighted various scenarios how misinformation are shared in social networking sites [6]. Henceforth, analysis of different tweets can be a helpful tool for the policy makers and health care organizations assess the needs of their stakeholders and address them appropriately. On Twitter, there is huge amount of data available on each and every possible topic or issues of the world. For which Twitter is also referred to as a gold mine for data collection. Further Twitter also provides its own REST (REpresentational State Transfer) API [33]. This API allows to access and retrieve Twitter data. In this paper, leveraging a set of tools (Twitter's search application programming interface (API)), Tweepy Python library and using a set of hashtags (“#Corona", “#Covid-19" etc.,), thousands of texts are extracted of public English language tweets for two different time frames (3/07/2020 to 11/07/2020 and 01/08/2020 to 06/08/2020). Tweepy contains variety of methods that can be used to access tweets through twitter application. There has been an exponential growth in the use of textual analytic, natural language processing (NLP) and use of artificial intelligence (AI) techniques in research and in the development of various applications. In the literature, various approaches like dictionary based, machine learning approaches are implemented by various researches for sentiment analysis of twitter data [33, 38, 26]. As a subsection, some of the latest available state-of-the-art literatures are highlighted below.

1.1 Literature survey

Before the COVID-19 pandemic there were several other viruses rapidly spread in different countries, Zika and swine flu are some of those viruses. Study of the sentiments of individuals on Twitter are useful to generate important health care information. This kind of study can be helpful in predicting the outbreak and its early detection. Grover *et al.* (2015) [32] have proposed a model named as named as Swine Flu Hint Algorithm (SEHA) to look after epidemic activities happen on Twitter related to swine flu influenza. Analysis was done to capture the real word dynamics with respect to the life cycle of the epidemic for which Markov chain model and Bag of words model were used. For forecasting Zika virus incidence, Sarah *et al.* (2017) [35] have conducted a study on twitter data and HealthMap digital surveillance system with historical Zika suspected case counts to track and predict estimates of suspected weekly Zika cases during the period of 2015 to 2016. Moving into COVID-19 issue, Alrazaq *et al.* (2020) [6] have analyzed English language tweets from February 2, 2020, to March 15, 2020. They have analyzed the collected tweets using word frequencies of single (uni-grams) and double words (bi-grams) model. Analysis identified 12 topics, which were grouped into four main themes: origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection. In another very recent work, public sentiment associated with the progress of Corona virus and COVID-19, Samuel *et al.* (2020) [34] have compare the effectiveness in classifying Corona virus Tweets of varying lengths. Naïve Bayes, logistic regression, linear regression and K-Nearest Neighbor (K-NN) models are evaluated for classifying the tweets. In another work authors have collected different tweets on COVID-19 from 11^{th} March 2020 to 31^{st} March 2020 [10]. Then after performing different preprocessing steps analyse how the citizens of 12 different countries are sharing their emotions and dealing with the situations. In a recent study, Singh *et al.* (2020) [37] has implemented the Susceptible-Infectious-Recovered (SIR) model based on Kermack-McKendrick theory of epidemic model to understand the outbreak of COVID-

19. This study attempt to estimate the number of people infected with a infectious disease in a closed population over time.

1.2 Motivations, objectives and contributions

After reviewing the related literature study, it is observed that different researchers have developed different models or perform suitable study towards the directions of detection of influenza like epidemic or other categories. As the world health care scenarios are changing every day, so to understand the current psychological behavior or opinion of people from social networking sites are also desirable. Also for performing text categorization using machine learning approaches, different types of learning algorithms are available in the literature. Which leads to the question of which learning algorithm is better. By keeping these observations this present study is undertaken with the following objectives: (i) to collect large amount of tweets for two different period of time; (ii) to categorize the tweets into negative, neutral and positive; (iii) to extract features from the tweets and to predict the tweets using different state-of-the-art inductive learning based classifier models; (iv) to perform a comprehensive comparisons between the learning models.

In order to satisfy the above mentioned objectives of this research, the research work is carry forward as follows: (i) Prepare different twitter datasets by authenticating and streaming real time search API provided by the Twitter web-site. Total 40,000 (20,000 + 20,000) tweets are collected in two different time frames.
(ii) Cleansing of the data is done to increase the performance of the analysis process.
(iii) By measuring the polarity and subjectivity of different tweets, tweets are categorized into *negative*, *neutral* and *positive* classes.
(iv) Term Frequency-Inverse Document Frequency (TF-IDF), feature extraction technique is used to transforming the text data into numerical features usable for learning models. Moreover, for further analysis, uni-gram, bi-gram and tri-gram models are applied to generate three different matrices for individual Tweeter datasets.
(v) Four different inductive learners namely, Gaussian Naïve Bayes (G-NB), Bernoulli's Naïve Bayes (B-NB), Random forest (RF), and Support vector machine (SVM) classifiers are used to train prediction models for further classifications of unseen tweets. Different evaluation measures are computed to compare the performances of the learning algorithms.

The key contributions of this research are as follows: (i) By measuring the score of the polarity and subjectivity of different tweets of two separate time frames, tweets are categorized into three classes. By counting the tweets of different categories, people opinions and assumption can be assess for the COVID-19 situations. (ii) Demonstration and comparisons of the performances of the G-NB, B-NB, RF, and SVM for Tweets classifications. For each dataset, three different N-grams models are used to prepare the TF-IDF feature extraction process. Then, the learning models are applied to check their performances with respect to accuracy, precision, recall, F1-Score and execution time.

1.3 Structure of the paper

The remainder of this paper is organized as follows. Section 2 describes the overall structure of the proposed methodology. The experimental setup and evaluations are highlighted in Section 3. Experimental result analysis are presented in Section 4. Thereafter, discussion and summarization of the study is highlighted in Section 5. The paper is ended with concluding remarks with future perspectives in Section 6.

2 Proposed Methodology

The pipeline of the methodology, which is followed in this research is shown as a block diagram in Figure 1. The proposed methodology is divided in to three phases viz., data collection, data preprocessing and data analysis. Before going into the detail steps, it is worthy to mention that Natural Language Processing Toolkit (NLTK), which is a python based platform is extensively used in this study [1, 2]. The data analysis phase is derived in two parts. The first part of the analytic includes sentiment analysis of the textual component of the Twitter data. Tweets are assigned sentiment scores of polarity and subjectivity by using different methods and Python libraries. The second part of analytic includes the feature extraction using TF-IDF technique for further prediction with inductive learning algorithms. Descriptions of the phases of the methodology are shown in Figure 1.

2.1 Data collection

Two sets of Corona virus related random tweets are collected between 3/07/2020 to 11/07/2020 (Dataset-I) and 01/08/2020 to 06/08/2020 (Dataset-II), using the Twitter standard search API using Python. Some predefined key terms, consisting of a set of the most widely used scientific and news media terms relating to the novel coronavirus are used. Text are extracted and stored in two different comma separated version (CSV) files. The tweeter texts and metadata of the tweets using the time

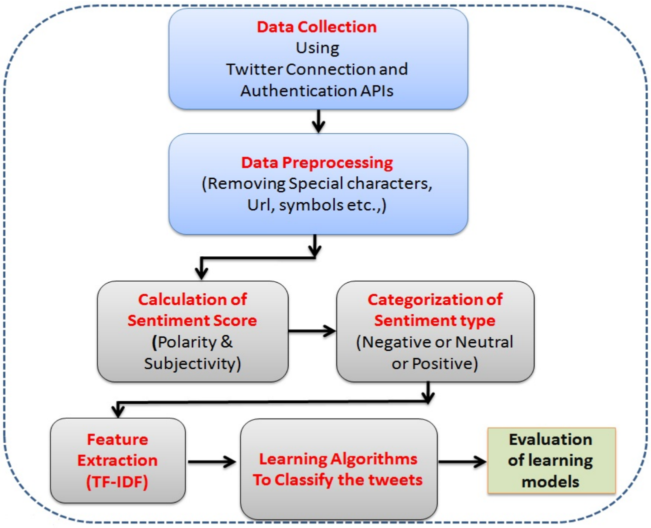


Figure 1: Block diagram of the proposed methodology with different phases.

stamp are stored. Only English language tweets were collected in this current study. The steps which are followed to scrap the twitter data are as follows:

Step 1: Create or register in Twitter Developer account on "developer.twitter.com". Thereafter, create an App on Twitter which will provide the access keys to access the twitter resources.

Step 2: The application keys namely "Consumer Key", "Consumer Secret", "Access token" and "Access Token Secret" will be used to access twitter resources and shall be used for authentication. Tweepy python is necessary to access the Twitter API.

Step 3: After the authentication step, with different hashtags #Covid19, #Crona, #Coronavirus, #CoronaIndia thousands of tweets are randomly downloaded and saved as a CSV file.

Total 40,000 tweets based on the aforementioned steps were collected.

However, during the process of data collection from twitter by scrapping through the python code few challenges were face and those includes: (i) Some tweets related to the above mentioned hashtags have relations with different images and video's also; (ii) though the language for the tweets considered as English, but some of the comments mix up with other languages also; (iii) lots of comments had unnecessary special characters, emoticons, urls, hashtags etc.

2.2 Data preprocessing

Pre-processing is an important phase in text processing. The data, which are extracted from the Twitter website is usually not clean and contains a lot of special characters, emoticons, urls, hashtags and other text which are not necessary for analysis purpose. In Figure 2, word cloud representation of the collected Tweeter data without performing any pre-processing steps is shown. Word cloud is one of the popular data visualization technique used for representing text data. Significant and frequent textual data can be highlighted using the word cloud representation. It is obvious from the Figure 2 that the maximum number of frequent words in the word cloud collections are the unused symbols or characters. Thus this word-cloud representation justify the significance of the pre-processing steps before performing any data analysis task. Henceforth, in order to perform the pre-processing task, in this work initially non-English tweets are identified using the language field in the tweets metadata and removed them to avoid further analysis. Thereafter, removed special characters, different non-printable characters from the tweets.

In Figure 3 (a), sample of Tweets without the pre-processing steps are shown for the two datasets. Similarly in Figure 3 (b), Tweets after the preprocessing steps are shown. It can be observed from those Figures that pre-processing step has cleaned the unnecessary characters from the tweets. To visualize the data properly word cloud representations of the two datasets are shown in Figure 4 (a)-(b). It shows commonly the most frequent words used in the cleaned Tweeter datasets. It can be observed from the word cloud visualizations that most of the dominated words are like: confirmed cases, covid, covid 19, plasma, mask, wear mask, pandemic, doctor, vaccine, clinic, health, treat, infected, recovery

Figure 2: Word cloud representation of Tweeter data without Pre-processing.

timestamp	tweet_text
03-07-2020 10:36	b'RT @WitsRHI: How can I reduce my risk of #Co...
03-07-2020 10:36	b'RT @GaetanBurgio: An update on #COVID19 situ...
03-07-2020 10:36	b'RT @ExeterMed: Missed Dr Bharat Pankhania @d...
03-07-2020 10:36	b'RT @ernestngBRO: If #covid19 was an anime PA...
03-07-2020 10:36	b'RT @DrTedros: We thank #Sweden for their gen...

timestamp	tweet_text
11-07-2020 04:43	b#CoronaInfoCH #COVID19 #corona people are cr...
11-07-2020 04:43	bRT @drharshvardhan: Ordered immediate replace...
11-07-2020 04:43	bRT @VABVOX: 1,000 kids in the hospital with #...
11-07-2020 04:43	bRT @QuickTake: \xe2\x9a\xa1\xef\xb8\x8f #Cov...
11-07-2020 04:43	bRT @GavinNewsom: WEEKEND REMINDER: #COVID19 i...

	timestamp	tweet_text
19995	2020-08-01 02:07:01	b'RT @TheLancetEndo: New #COVID19 Corresponden...
19996	2020-08-01 02:07:01	b'Doing my part to help teach my grandson to d...
19997	2020-08-01 02:07:00	b'RT @TelanganaCMO: CM Sri KCR conveyed Id-Ul-...
19998	2020-08-01 02:07:00	b'The impact of the pandemic on the well-being...
19999	2020-08-01 02:06:59	b'RT @ConservaMomUSA: DESPICABLE- @RepMaxineWa...

(a)

timestamp	tweet_text
03-07-2020 10:36	How can I reduce my risk of Covid19 infection ...
03-07-2020 10:36	An update on COVID19 situation in Australia. T...
03-07-2020 10:36	Missed Dr Bharat Pankhania talking aout the C...
03-07-2020 10:36	If covid19 was an anime PA13 (1/3) We
03-07-2020 10:36	We thank Sweden for their generous and ongoing...

timestamp	tweet_text
11-07-2020 04:43	CoronaInfoCH COVID19 corona ireland trendingne...
11-07-2020 04:43	Ordered immediate replacement of Medical Super...
11-07-2020 04:43	1,000 kids in the hospital with COVID19 in Tex...
11-07-2020 04:43	Covid19 can spread in through the air indoor...
11-07-2020 04:43	WEEKEND REMINDER COVID19 is not over just ecau...

19993	2020-08-01 02:07:02	citizen If COVID19 didnt get you ScottyFromMar...
19994	2020-08-01 02:07:01	politicizes HermanCains death with empty con...
19995	2020-08-01 02:07:01	New COVID19 CorrespondencenSARS-CoV--related a...
19996	2020-08-01 02:07:01	Doing my part to help teach my grandson to do ...
19997	2020-08-01 02:07:00	CM Sri KCR conveyed Id-Ul-Zuha greetings to th...
19998	2020-08-01 02:07:00	The impact of the pandemic on the well-eing of...
19999	2020-08-01 02:06:59	politicizes HermanCains death with empty con...

(b)

Figure 3: (a) Without any pre-processing Sample of Tweets from (b) After pre-processing sample of Tweets.

etc. Hence, these word cloud representation highlights the most significant words, it is not referring to any unnecessary words. So the benefits of the pre-processing steps and cleaning of the large quantity of tweets can be clearly identified. Moreover, such high used words also useful for inclining to generate hints about the COVID-19 epidemic.

It is also necessary to count the number words in the tweets. Because very less numbers of words in a tweet may not give required information or it may mislead also. Therefore, the numbers of word counts in each tweets are calculated. In Figure 5 (a)-(c), frequency of word counts are shown.

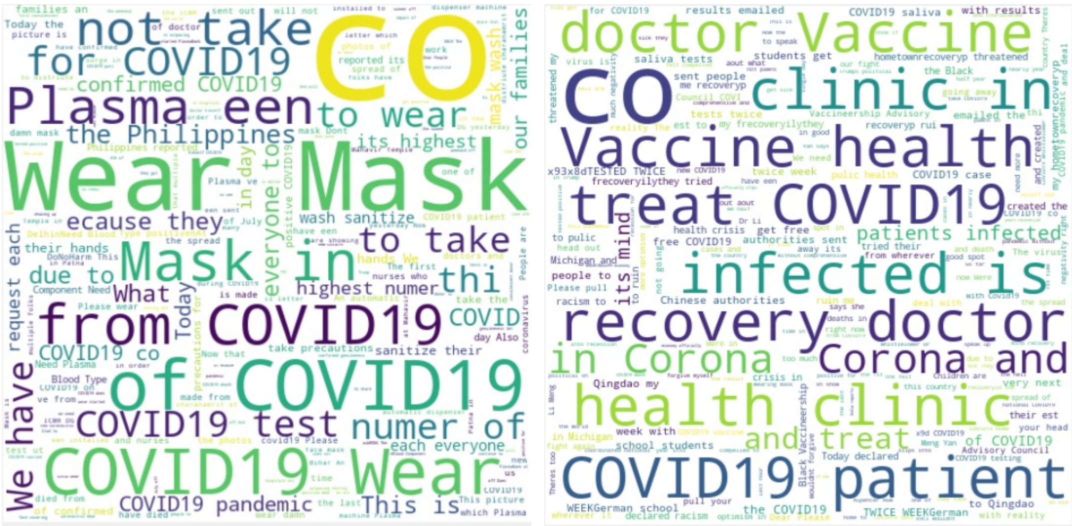
2.3 Subjectivity and polarity

After cleaning of the tweets, using the NLTK and TextBlob libraries the polarity and subjectivity of each tweets are calculated from both the CSV files. To be precise, sentiment analysis is basically the process of determining the attitude or the emotion of the writer, i.e., whether the sentence emotions inclined towards *negative* or *neutral* or *positive* directions. The 'sentiment' method of TextBlob returns two properties, polarity, and subjectivity [7, 21].

Definition 1(Sentence Subjectivity): An objective sentence expresses some factual information about the world, while a subjective sentence expresses some personal feelings or beliefs. e.g., "She was quite happy with her earlier days." Subjective expressions may come in many forms, e.g., opinions, allegations, desires, beliefs, suspicions, and speculations [21].

Definition 2(Sentence Polarity): Polarity refers to identifying sentiment orientation i.e., *negative*, *neutral* and *positive* in written or spoken language.

Polarity is a float value which lies in the range of [-1, 1]. The polarity 1 means positive statement and -1 means a negative statement and 0 means neutral. Sub-



(a) (b)

Figure 4: Word cloud representations after pre-processing (a) Dataset-I (b) Dataset-II,

jective sentences generally refer to personal opinion, emotion or judgment. Subjectivity score is also a float value which lies in the range of [0, 1]. When it is close to 0, it is more about facts. When subjectivity increases, it comes close to be an opinion.

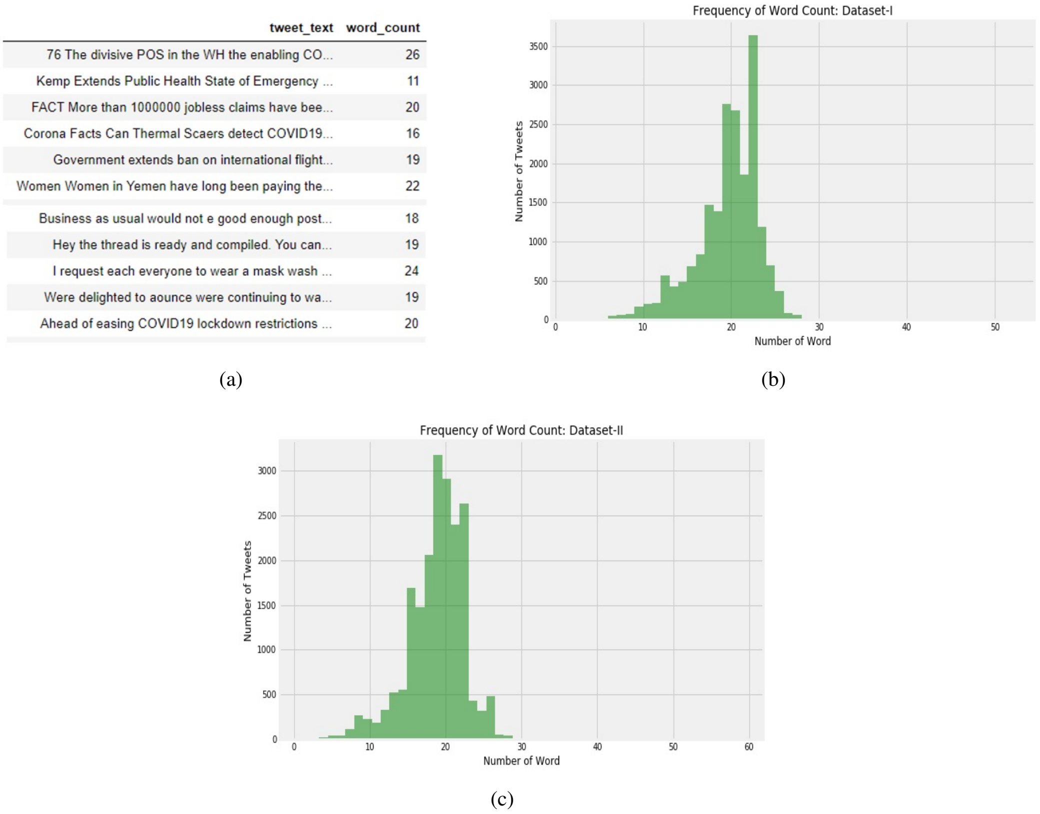
Say for instance, for the above statement "She was quite happy with her earlier days", lets calculate the polarity and subjectivity using the *sentiment* function of Textblob python library. Say the polarity is 0.5, and subjectivity is 0.58; then these values says that the statement is positive and mostly it is a public opinion of somebody, it can not be a factual information. So as per the values of subjectivity and polarity, sentiment is decided by using TextBlob.

In Figure 6 (a) and Figure 6 (b), histogram representation of the polarity scores for the two different datasets are shown. It can be observed from the histogram representation that the scores are not distributed properly. Maximum peaks are near the ranges of from (0 - 0.25). Similarly the distribution of the subjectivity score for the two different datasets are shown in Figure 7 (a) and Figure 7 (b). Then in Figure 8 for the datset-II, a snapshot of the sample results of the tweets with their subjectivity-score, polarity-score are shown. Based on the values categorization of the tweets are performed. If the polarity is above 0 it is categorized as *positive*, if it is 0 it categorized as *neutral* and if the polarity score is below 0 it considered as *negative*.

2.4 Term frequency-inverse document frequency (TF-IDF)

To apply any prediction model it is required to transform the Tweeter text data into numbers. This process is known as text vectorization. It is a fundamental step in the process of machine learning for analyzing text. It is also to be noted that different vectorization approach may drastically affect the end of the results. In the literature, various feature extraction techniques are available such as: bag of words model, word2vec, doc2vec etc [8]. But by looking at the computational effort and observing successful results [14, 18], in this study Term frequency-inverse document frequency (TF-IDF) technique is adopted.

TF-IDF technique is one of the popular approach used in information retrieval and text mining for doing the text vectorization or extraction. TF-IDF uses two statistical methods, first is Term Frequency (TF) and the other is Inverse Document Frequency (IDF). TF refers to the total number of times a given word (term) t appears in a document (tweet) against the total number of all words in the document. On the other hand, IDF measure of how unique the words are i.e., it means, how common or rare a word is in the entire document set. Tweets can be consider as a shorter document. The product of TF and IDF provides a measure of how frequent the word is in a document multiplied by how unique the word is, and that is the TF-IDF measure. The words that are common in every tweets, such as: "this", "what", or say "if", get low rank, even



	tweet_text	word_count
76	The divisive POS in the WH the enabling CO...	26
	Kemp Extends Public Health State of Emergency ...	11
	FACT More than 1000000 jobless claims have bee...	20
	Corona Facts Can Thermal Scaers detect COVID19...	16
	Government extends ban on international flight...	19
	Women Women in Yemen have long been paying the...	22
	Business as usual would not e good enough post...	18
	Hey the thread is ready and compiled. You can...	19
	I request each everyone to wear a mask wash ...	24
	Were delighted to aounce were continuing to wa...	19
	Ahead of easing COVID19 lockdown restrictions ...	20

Figure 5: (a) Sample of frequency of word counts (b) histogram plot of word count in Dataset-I (c) histogram plot of word count in Dataset-II

though these words may appear many times, since they don't mean much to that tweet sentences.

To put it in more formal mathematical terms, the TF-IDF score for a word t in the short document sd from the document set D is calculated as follows:

$$tf_{t,sd} = \frac{n_{t,sd}}{T} \quad (1)$$

Here, in Eq. (1) in the numerator, n is the number of times the term t appears in the document sd.

$$idf_t = log\frac{N}{M} \quad (2)$$

Here, in Eq. (2) in the numerator, N is the total number of short documents (tweets) and M is the number of tweets with term t.

Multiplying these two measures, results in the TF-IDF score of a word in a tweet. The higher the score, the more relevant that word appears in that particular tweet.

For instance, if we take our set of sentences as: T1 = "When Indias corona case is at peak." and T2 = "Peak corona case in India."

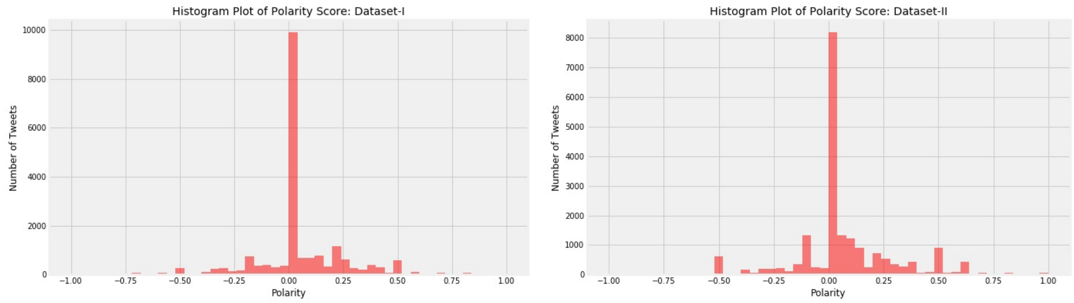
By applying a bi-gram approach [22] it will create a vocabulary set like the following: V = ["When Indias", "Indias corona", "corona case", "case is", "is at", "at peak", "Peak corona"... "in India"]

V = ['at peak', 'case in', 'case is', 'corona case', 'in india', 'indias corona', 'is at', 'peak corona', 'when indias']
The TF values:
T1 = [1 0 1 1 0 1 1 0 1]
T2 = [0 1 0 1 1 0 0 1 0]

TF-IDF weighted:
T1 = [1.40546511, 0, 1.40546511, 1, 0,1.40546511, 1.40546511, 0, 1.40546511]
T2 = [0, 1.40546511, 0, 1, 1.40546511, 0, 0, 1.40546511, 0]

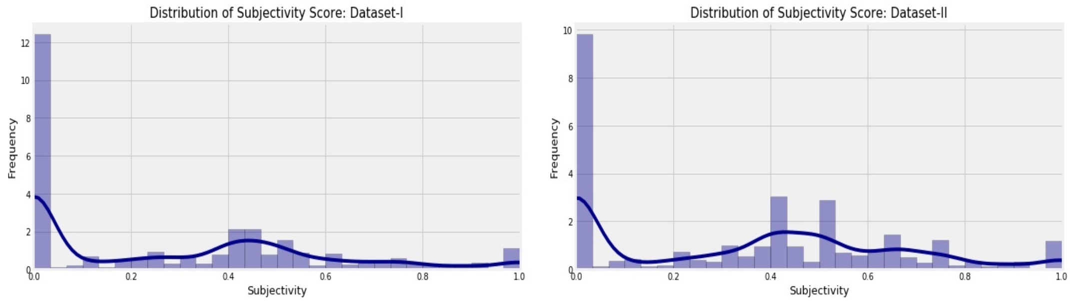
2.5 Inductive learning models for Classification

An inductive learning model is a (non)parametric model structure plus a criterion for tuning its degrees of free-



(a) (b)

Figure 6: Histogram of the Polarity score (a) Dataset-I (b) Dataset-II



(a) (b)

Figure 7: Distribution of subjectivity score (a) Dataset-I (b) Dataset-II

dom over experimental data produced by some unknown phenomenon [5, 23, 11]. So, the problem of induction, is how to draw general conclusions about the future from specific observations from the past. So, formally given a set of instance data (x) there is predefined output say y or y in the form of a function $f(x)$. The goal of the inductive learning is to learn the function from the given set of examples/data for a new unseen data say $x^{'}$. Thus, in this present study for each tweet, the output is considered from the output set *negative* or *neutral* or *positive*. For both the datasets, each tweets (20,000 and 20,000) output labels are assigned. To predict the unseen tweets the hypothesis/function needs to be learned by the inductive models from a given input set of data.

The selection of a suitable inductive learning method is a crucial phase in any system design methodology. Here in this study, by observing various previous works [7, 34, 19, 25, 24] related to text classification and tweeter data classification, three different inductive learning models as the classifiers are considered viz., Naïve Bayes(NB), Random Forest (RF), and Support Vector Machine (SVM). Two variations of the NB model are used in this study, namely Gaussian- NB (G-NB), and Bernoulli's- NB (B-NB). The overall process of the data analysis process is shown in Figure 9.

NB is a well known simple and effective method for text classification [34, 19]. It has been used widely for document classification since 1960s. This classifier is theoretically based on the Bayes theorem. NB uses maximum posteriori estimation to find out the class (i.e., features are assigned to a class based on the highest conditional probability). Bernoulli NB is useful when the features representations are binary and Gaussian NB it is useful when working with large continuous values. By observing various previous study in this research Gaussian-NB, Bernoulli-NB models are considered.

RF classifier is basically a bagging technique which falls under the ensemble learning [25, 17] techniques. It is a combination of different decision trees which are considered as the base learners. Each decision trees are trained with feature and row sampling with replacement concept with the same distribution. Thus the trees are trained independently and an unknown samples are labeled according to the majority vote rule: i.e., it is la-

	timestamp	tweet_text	Subjectivity	Polarity	Analysis
	....	....	....		
19990	2020-08-01 02:07:03	Just a snipppet of drile from this morning......	0.333333	0.250000	Positive
19991	2020-08-01 02:07:02	Guys this is my super protective Italian style...	0.555556	0.444444	Positive
19992	2020-08-01 02:07:02	76 The divisive POS in the WH the enaling COM...	0.500000	0.625000	Positive
19993	2020-08-01 02:07:02	citizen If COVID19 didnt get you ScottyFromMar...	0.400000	-0.100000	Negative
19994	2020-08-01 02:07:01	politicizes HermanCains death with empty con...	0.500000	-0.100000	Negative
19995	2020-08-01 02:07:01	New COVID19 CorrespondencenSARS-CoV--related a...	0.327273	0.068182	Positive
19996	2020-08-01 02:07:01	Doing my part to help teach my grandson to do ...	0.535714	0.285714	Positive
19997	2020-08-01 02:07:00	CM Sri KCR conveyed Id-Ul-Zuha greetings to th...	0.000000	0.000000	Neutral
19998	2020-08-01 02:07:00	The impact of the pandemic on the well-eing of...	0.383333	0.116667	Positive
19999	2020-08-01 02:06:59	politicizes HermanCains death with empty con...	0.500000	-0.100000	Negative

20000 rows × 5 columns

Figure 8: Sample results of the tweets with their subjectivity-score, polarity-score and categorization into three classes *negative* or *neutral* or *positive*.

beled with the most popular class among those provided by the ensemble trees. The ensemble concepts of different decision trees in RF classifier leads the system towards low bias and low variance.

SVM is a supervised learning algorithm which is widely used for the classification task [25, 24]. SVM is based on the idea of finding a hyperplane that best separates the features into different domains. SVM is applicable for the data which are linearly separable and also it applicable non linear data with proper kernel functions or tricks. The data are mapped through a Gaussian/radial basis function or other type of kernel (linear, polynomial) tricks to a feature space in a larger dimension space with the aim being maximum separation between classes. In this research linear kernal trick function is applied.

The two datasets (Dataset-I & Datset-II) are sliced into a training set and test set. Training set is the subset of the overall dataset to train the learning model. On the other hand test set is the subset to test the trained model. For both the datasets 70% of the tweets are used as the train set and 30% of the data are considered as the test set.

3 Experimental Setup and Evaluation

In this section, the details of the experimental evaluation starting with the experimental setup, followed by the quantitative measures used for this research are demonstrate. For the sake of uniform experimental results, all the methods are implemented in Python. Programs are simulated in a machine with Processor: IntelÂ® Xenon(R) CPU- E5-1630, 3.70 GHz clock speed, main memory of 32GB and having Windows 10 environment.

The detailed experimental setup are as follows:

1. All the methods and functions are implemented in Python 3.7 environment. From the implementation point of view, data structures like data frames, dynamic lists, collections, and dynamic arrays are used.

2. The partitioned of the individual datasets are performed according to a train-test (70% and 30%) spilt scheme. For RF classifier, number of estimator i.e., the number of decision trees in the forest is set to 100 and the maximum depth of the tree is set 'none'. For the SVM classifier, parameters like kernal is set to 'linear' and the multi-class support is handled according to a one-vs-one scheme. For the Gaussian NB and Bernoulli NB classifiers default setting of the 'sklearn' environment is used.

The evaluation measures for multi-class classification problems are highlighted in Table 1 [39]. Here the classification problem is related multi-class because the classifiers have to deal with three classes viz., negative, neutral and positive. For a particular class C_i

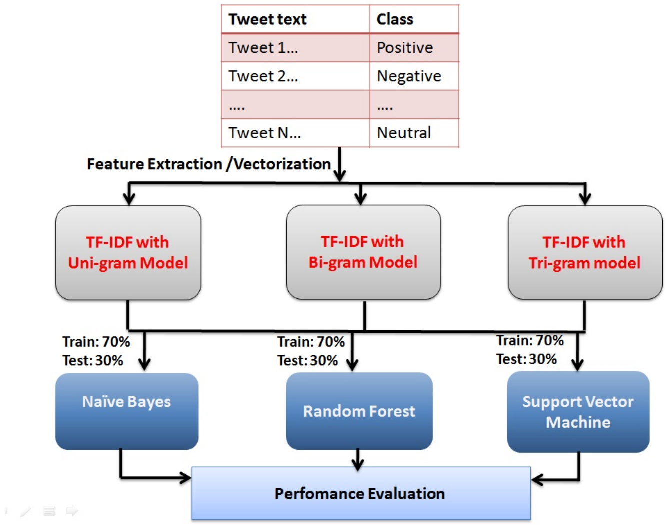


Figure 9: A block diagram of the overall learning process.

the assessment is defined by TP_i, TN_i, FP_i, FN_i are calculated from the counts for C_i. The quality of the overall classification results are assessed in three ways: first measure is the average of the same measures calculated for $C_1, C_2 ... C_L$ i.e., the macro-averaging which is shown with M indices. Second is the sum of counts to obtain cumulative TP, FN, TN, FP and then calculating a performance measure i.e., the micro-averaging which is shown with μ indices. Macro-averaging treats all classes equally while micro-averaging favors bigger classes. The third one is the weighted averaging, with working example it is highlighted in the Sub-section 4.2.

4 Experimental Results Analysis

4.1 Analysis of the boxenplots, scatter plots and bar charts

The Boxenplot (or "letter-value-plot")[15] has some similarities with box plot but it is much more useful. As the polarity value may be long tailed, by using this plot different quartile of polarity values can be shown. Thus, by plotting several quartile values, it enables to understand the shape of the distribution particularly in the head end and in the tail end. The inner-most box represents 25-75% inner quartile range and bold line is the median line. In the Figure 10 (a) the boxenplot for Dataset-I is shown. It can be observed that *negative* category has long tailed polarity compared to *positive* tweets. Similarly in Figure 11 (a) the boxenplot for Datset-II is shown. Here it is interesting notice that category *positive* has higher polarity compared to *negative* tweets.

In the Figure 10 (b) scatter plot of the sentiment of each tweet based on the polarity and subjectivity of the tweets are shown. The maximum density of the positive tweets, for the polarity value range up to 0.5 and the subjectivity ranges from 0.2 to 0.8. When the subjectivity range is 0.8-1.00 and polarity is 0.75-1.00, the number of positive tweets are less. The red dotted points represents the neutral tweets. Similarly for dataset-II in Figure 11 (b) scatter plot is shown.

The percentage of negative, neutral, and positive tweets for the Dataset-I are 18.10%, 48.50%, and 33.40%. Similarly the percentage of negative, neutral, and positive tweets for the Dataset-II are 19.20%, 38.90%, and 41.90%. As a bar chart the numbers of tweets belongs to each categories are shown in Figure 10 (c) and Figure 11 (c) for both the datasets. It can be observed that the numbers of positive tweets are high in the second dataset. Where else in the first set of dataset the neutral tweets are high.

4.2 Analysis of classifiers evaluation measures

In order to predict the categories of tweets into negative, neutral, and positive, experiments are carried out using train set/test set spilt applied on the four popular classifiers viz., Gaussian-NB, Bernoulli-NB, RF, and SVM classifiers. The overall summary of the experimental results achieved on the two datasets (Dataset-I and Dataset-II) by using 70% training and 30% test samples are reported in Table 2 and Table 3. In Figure 12 (a)-(c) and Figure 13 (a)-(c), the best classification

Table 1: Evaluation measures for the multi-class classification of Twitter data

Measure	Formula	Description
Avg. Accuracy (Acc)	$\frac{\sum_{i=1}^{L}\frac{TP_i+TN_i}{TP_i+FN_i+TN_i+FP_i}}{L}$	The average per-class effectiveness of a classifier.
Precision (Pre)	$\frac{TP_i}{TP_i+FP_i}$	Number of correctly classified positive examples divided by the total number of examples that are classified as positive.
Recall (Rec)	$\frac{TP_i}{TP_i+FN_i}$	Number of correctly classified positive examples divided by the number of true positives plus the number of false negatives.
F1-Score	$2 \times \frac{Precision \times Recall}{Precision+Recall}$	It is the harmonic mean of precision and recall.
μ-Avg.Precision	$\frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L} TP_i+FP_i}$	Agreement of the data class labels with those of a classifiers if calculated from sums of per-text decisions.
μ-Avg.Recall	$\frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L} TP_i+FN_i}$	Effectiveness of a classifier to identify class labels if calculated from sums of per-text decisions.
μ-Avg.F1-Score	$2 \times \frac{\mu-Avg.Precision \times \mu-Avg.Recall}{\mu-Avg.Precision+\mu-Avg.Recall}$	Relations between data positive labels and those given by a classifier based on sums of per-text decisions.
M-Avg.Precision	$\frac{\frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L} TP_i+FP_i}}{L}$	Average the per-class agreement of the class labels with those of a classifier.
M-Avg.Recall	$\frac{\frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L} TP_i+FN_i}}{L}$	Average per-class effectiveness of a classifier to identify class labels.
M-Avg.F1-Score	$2 \times \frac{M-Avg.Precision \times M-Avg.Recall}{M-Avg.Precision+M-Avg.Recall}$	Relations between positive labels and those given by a classifier based on a per-class average.

evaluation results achieved by the different classifiers on uni-gram, bi-gram and tri-gram models applied on Dataset-I and Dataset-II are shown.

For instance, let us consider the confusion matrix which is shown in Figure 12 (c) and compute the μ-Avg. Recall value. As there are three classes viz., *negative*, *neutral* and *positive* the TP and FP values with respect to all the classes are as follows:

$TP_{negative}$ = 866; $TP_{neutral}$ = 2940; $TP_{positive}$ = 1398;

$TP_{negative}$ = (1 + 6) = 7; $TP_{negative}$ = (211 + 565) = 776; $TP_{negative}$ = (4 + 9) = 13;

Thus, the μ-Avg. Recall = $\frac{866+2940+1398}{866+2940+1398+7+776+13} = \frac{5204}{6000}$ = 0.867 = 0.87

Similarly, let's compute the 'Weighted Average precision' (W-Avg. Precision). It is computed as follows:

W-Avg. Precision = Actual number of Class *negative* instance (tweet) $\times$ Recall of Class *negative* + Actual number of Class *neutral* instance (tweet) $\times$ Recall of Class *neutral* + Actual number of Class *positive* instance (tweet) $\times$ Recall of Class *positive*

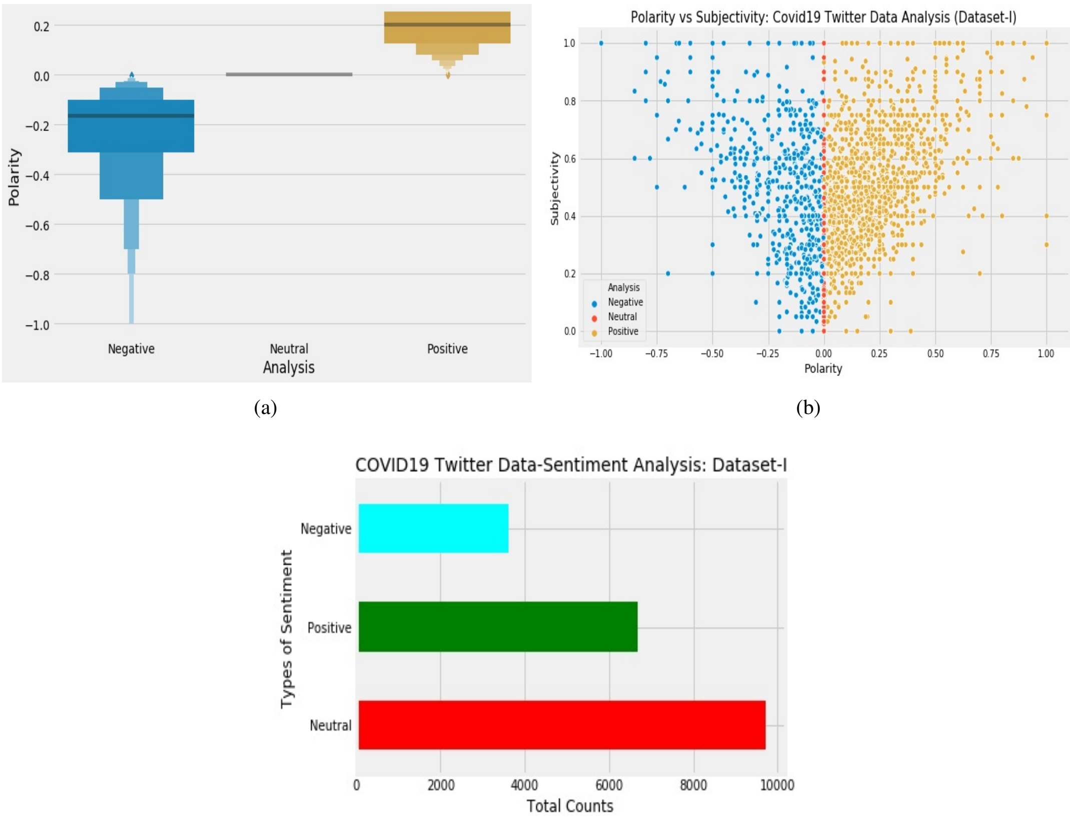
= $\frac{1081}{6000} \times 0.99 + \frac{2950}{6000} \times 0.79 + \frac{1969}{6000} \times 0.99$

= 0.178 + 0.387 + 0.324

= 0.889 = 0.89

Now, it can be observed from the summarized experimental results shown in Table 2 and Table 3 that RF classifier has achieved better classification evaluation results with bi-gram and tri-gram models compared to SVM and the NB classifiers (Gaussian and Bernoulli). However, SVM has achieved slight better classification accuracy of 1.46% compared to RF model with respect to uni-gram model. But the execution time comparison graph which is shown in Figure 14 (a), implies that SVM has spend a larger amount of time. For the other two models also i.e., for bi-gram and tri-gram, the execution time of SVM is very high compared to other three learning models. It interesting to notice that Bernoulli NB has very lesser amount of computational time with respect to all the N-gram models and in comparison to other learning models. But the accuracy achieved by this model is comparatively bit lesser (0.97% and 0.07%) than the RF classifier in bi-gram and tri-gram model. For tri-gram model the amount of classification accuracy difference between RF and Bernoulli NB is negligible. Moreover, Bernoulli NB has better computational cost.

Similarly it is noticeable from Table 3 that SVM classifier has achieved better classification evaluation results with respect to uni-gram model, compared to other learning models. But the computational cost for SVM is very high, which can be observed from Figure 14 (b). Like the Dataset-I, in this Dataset-II also RF achieved better results with respect to bi-gram & tri-gram models. But if we observe the results of Bernoulli's NB than it has very less amount of computational cost, and the classification accuracy is bit lesser than RF.



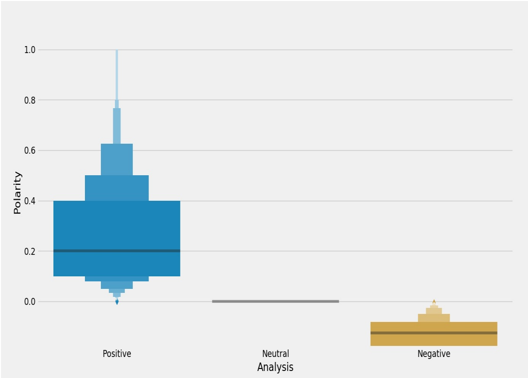
(a) (b)

(c)

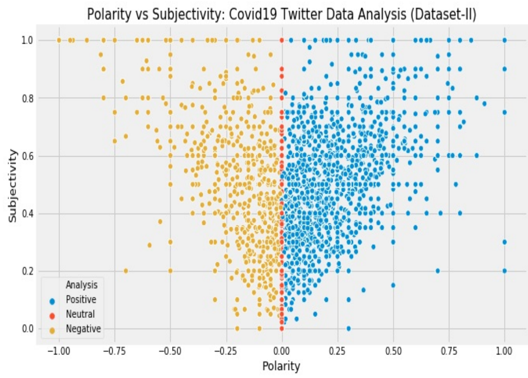
Figure 10: Graphical representation (a) Boxenplot for polarity score (b) Scatter plot of Sentiment Polarity vs Subjectivity (c) Bar chart for the count of Positive, Negative & Neutral Tweets of Dataset-I

5 Discussions

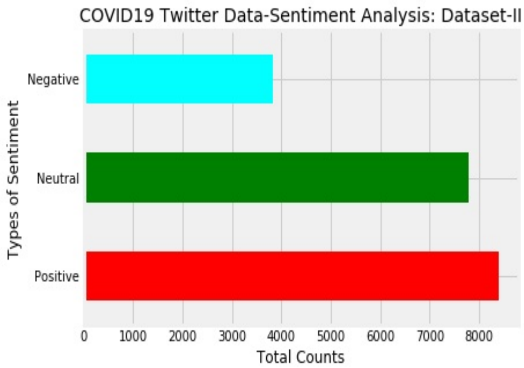
Textual analytic has gained significant attention over the past few years with the advent of technological improvements. In this research, attempt is made to perform the sentiment analysis on the ongoing COVID-19 pandemic related public sentiments of Twitter data. This study also intended to explore the viability of the inductive learning models for the predictions of unseen tweets. Initial analysis was done to categorize the two different sets of datasets collected in two separate time frames into three distinct categorize (viz., *negative*, *neutral* and *positive*) using the NLTK framework with the assistance of the python libararies. From that categorization, it is observed that the number of *positive* tweets in the second set of dataset is significantly higher than the *neutral* and *negative* tweets. Which shows a positive directions towards the individual behaviors on this pandemic situation. But it is also worthy to be mentioned that the Twitter environment is totally dynamic. By only taking two sets of limited numbers of Twitter datasets it is difficult to come up to a conclusive evidence that the individual opinions for this pandemic is going towards a positive directions, but this short study maybe helpful to generate the hint about the epidemics. Also, for the textual analytics different visualization plots/tools are used to represent the data for better understanding and for visibility. In the second phase of the data analysis part, TF-IDF based feature extraction and inductive learning models are applied. Though in this study formally the classification models are not stated properly but the experimental comparisons of the inductive learning classification models, gives sufficient directional support for the use of RF model and B-NB model for the prediction/classifying unseen twitter data into any of the three different categories. The extensive experiments with different classifiers evaluation measures suggests that RF model has performed significantly better compared to other three models. The execution time study justifies that the use of Linear-SVM is computationally very expensive. However, the uses



(a)



(b)



(c)

Figure 11: Graphical representation (a) Boxenplot for polarity score (b) Scatter plot of Sentiment Polarity vs Subjectivity (c) Bar chart for the count of Positive, Negative & Neutral Tweets of Dataset-II

of B-NB with respect to execution time and with little compromise of evaluation measures performance can be a good choice to structure a computational model for text categorization.

5.1 Limitations and scope for improvements

Twitter data alone may not be sufficient to reflect the general mass sentiments for a nation or for the states. However, this research may provide a clear indication or direction for other comprehensive analysis of multiple textual data sources. The data sources may includes: Facebook data, different news articles and some other personal communications data. In this study, only English language based Twitter textual information are considered. However, there is high scope to incorporate other different languages for better performance of the model. Also, there is scope to apply the feature selection [9] process before the classification phase.

6 Concluding Remarks and Future Work

The propagation of data produced by the social networking sites spawns an amount of practical problems to be answered and analysis of sentiments is one of them. In this paper, by keeping in view of the ongoing COVID-19 pandemic, an attempt is made to analyse the performance of classical inductive learning models of classifiers to predict different tweets into three different classes. For doing the same, extensive experiments are conducted in two phases. In the initial phase, tweets were collected from Twitter site between 3/07/2020 to 11/07/2020 and 01/08/2020 to 06/08/2020, using Tweepy python library. Thereafter, with the assistance of Python based libraries and generated functions data are preprocessed. Word cloud visualization justifies the act of pre-processing steps. Then by calculating the polarity and subjectivity measures of the tweets, *negative*, *neutral* and *positive* are categories are created. In the later phase, well known TF-IDF feature extraction scheme with the uni-gram, bi-gram, and tri-gram

Table 2: Summary of experimental results of different classifiers evaluation measures on 70% training and 30% test set obtained by G-NB, B-NB, RF, and Linear-SVM classifiers on Dataset-I.

	Uni-gram					Bi-gram					Tri-gram				
Methods	**-**	**Pre**	**Rec**	**F1-Sc**	**Acc**	**-**	**Pre**	**Rec**	**F1-Sc**	**Acc**	**-**	**Pre**	**Rec**	**F1-Sc**	**Acc**
G-NB	Neg	0.75	0.86	0.80		Neg	0.80	0.86	0.83		Neg	0.60	0.91	0.72	
	Neu	0.89	0.89	0.73		Nue	0.92	0.87	0.90		Nue	0.94	0.84	0.89	
	Pos	0.62	0.62	0.72		Pos	0.84	0.88	0.86		Pos	0.91	0.78	0.84	
	μ-Avg.	0.74	0.74	0.74	73.70	μ-Avg.	0.87	0.87	0.87	86.33	μ-Avg.	0.83	0.83	0.83	83.71
	M-Avg.	0.75	0.75	0.75		M-Avg.	0.85	0.87	0.86		M-Avg.	0.81	0.85	0.82	
	W-Avg.	0.77	0.77	0.74		W-Avg.	0.87	0.87	0.87		W-Avg.	0.87	0.83	0.84	
B-NB	Neg	0.87	0.82	0.84		Neg	0.98	0.77	0.86		Neg	1.00	0.76	0.86	
	Neu	0.96	0.84	0.89		Neu	0.90	0.92	0.91		Neu	0.82	0.97	0.89	
	Pos	0.77	0.94	0.85		Pos	0.83	0.90	0.86		Pos	0.90	0.78	0.83	
	μ-Avg.	0.87	0.87	0.87	86.75	μ-Avg.	0.88	0.88	0.88	88.33	μ-Avg.	0.82	0.87	0.87	86.66
	M-Avg.	0.87	0.86	0.86		M-Avg.	0.90	0.86	0.88		M-Avg.	0.90	0.83	0.86	
	W-Avg.	0.88	0.87	0.87		W-Avg.	0.89	0.88	0.88		W-Avg.	0.88	0.87	0.86	
RF	Neg	0.98	0.85	0.91		Neg	0.99	0.83	0.90		Neg	0.99	0.80	0.89	
	Neu	0.90	0.99	0.94		Neu	0.83	0.99	0.91		Neu	0.79	1.00	0.88	
	Pos	0.97	0.89	0.93		Pos	0.97	0.78	0.87		Pos	0.99	0.71	0.83	
	μ-Avg.	0.93	0.93	0.93	93.40	μ-Avg.	0.89	0.89	0.89	**89.30**	μ-Avg.	0.87	0.87	0.87	**86.73**
	M-Avg.	0.95	0.91	0.93		M-Avg.	0.93	0.86	0.88		M-Avg.	0.92	0.84	0.87	
	W-Avg.	0.94	0.93	0.93		W-Avg.	0.91	0.88	0.88		W-Avg.	0.89	0.87	0.86	
SVM	Neg	0.96	0.91	0.93		Neg	0.99	0.82	0.90		Neg	0.99	0.79	0.88	
	Neu	0.94	0.98	0.96		Neu	0.82	0.98	0.89		Neu	0.79	1.00	0.88	
	Pos	0.95	0.93	0.94		Pos	0.95	0.76	0.84		Pos	0.99	0.70	0.82	
	μ-Avg.	0.95	0.95	0.95	**94.86**	μ-Avg.	0.88	0.88	0.84	87.80	μ-Avg.	0.86	0.88	0.86	86.38
	M-Avg.	0.95	0.94	0.94		M-Avg.	0.92	0.85	0.88		M-Avg.	0.92	0.83	0.86	
	W-Avg.	0.95	0.95	0.95		W-Avg.	0.89	0.88	0.88		W-Avg.	0.89	0.86	0.86	

techniques different features are extracted to prepare the datasets to feed it into the prediction models. Out of the four prediction models, experimental results suggest that RF and B-NB models performance are better than the other two state-of-the-art classifier models. Experimental study also shows that linear-SVM has higher computational cost. Moreover, this kinds of study for doing the analysis on pandemic issues can be useful to build better models which can predict the epidemic eruption and its outlines.

In this study presently English language based Twitter textual information are measured, author is presently working towards incorporating other languages. Also, the identification of fake and misinformation accompanied with this disease is highly desirable.

References

[1] Natural language toolkit (NLTK). `http://www.nltk.org/`.

[2] scikit-learn machine learning in python. `https://scikit-learn.org/stable/`.

[3] World health organization corrona virus dashboard. `https://covid19.who.int/`.

[4] A. B. Xie, O. R. R. P., I. Vovsha. Sentiment analysis of twitter data. *In Proceedings of the workshop on Languages in Social Media, Columbia University, New York*, pages 30–38, 2011.

[5] Alippi, C. and Braione, P. Classification methods and inductive learning rules: What we may learn from theory. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews*, 36(5):649–655, 2006.

[6] Alrazaq, A. A., Alhuwail, D., Househ, M., Hamdi, M., and Shah, . Top concerns of tweeters during the covid-19 pandemic: Infoveillance study. *Journal of Medical Internet Research*, 22(4):1–9, 2020.

[7] Ankit and Saleena, N. An ensemble classification system for twitter sentiment analysis. *Procedia Computer Science*, 132(2):937–946, 2018.

[8] Avinash, M. and Sivasankar, E. A study of feature extraction techniques for sentiment analysis. *Arxiv:1906.01573*, pages 1–12, 2019.

[9] Bania, R. Survey on feature selection for data reduction. *International Journal of Computer Applications*, 94(18):1–7, 2014.

[10] Dubey, A. D. Twitter sentiment analysis during covid-19 outbreak. *https://ssrn.com/abstract=3572023 or http://dx.doi.org/10.2139/ssrn.357202*, pages 1–9, 2020.

[11] Dumais, S., Platt, J., Heckerman, D., and Sahami, M. Inductive learning algorithms and represen-

Table 3: The summary of results in terms of different classifiers evaluation measures on 70% training and 30% test set obtained by G-NB, B-NB, RF, and Linear-SVM classifiers on Dataset-II.

	Uni-gram					Bi-gram					Tri-gram				
Methods	**-**	**Pre**	**Rec**	**F1-Sc**	**Acc**	**-**	**Pre**	**Rec**	**F1-Sc**	**Acc**	**-**	**Pre**	**Rec**	**F1-Sc**	**Acc**
G-NB	Neg	0.74	0.84	0.79		Neg	0.77	0.82	0.80		Neg	0.58	0.89	0.70	
	Neu	0.88	0.74	0.80		Neu	0.89	0.82	0.85		Neu	0.91	0.78	0.84	
	Pos	0.79	0.87	0.83		Pos	0.86	0.89	0.87		Pos	0.92	0.82	0.87	
	μ-Avg.	0.81	0.81	0.81	81.1	μ-Avg.	0.85	0.85	0.85	85.10	μ-Avg.	0.82	0.82	0.82	82.10
	M-Avg.	0.80	0.82	0.81		M-Avg.	0.84	0.85	0.84		M-Avg.	0.80	0.83	0.81	
	W-Avg.	0.82	0.81	0.81		W-Avg.	0.85	0.85	0.85		W-Avg.	0.85	0.82	0.83	
B-NB	Neg	0.92	0.78	0.84		Neg	0.99	0.75	0.85		Neg	1.00	0.73	0.85	
	Neu	0.88	0.92	0.90		Neu	0.82	0.95	0.88		Neu	0.76	0.98	0.86	
	Pos	0.89	0.91	0.90		Pos	0.92	0.87	0.89		Pos	0.96	0.80	0.87	
	μ-Avg.	0.89	0.89	0.89	86.75	μ-Avg.	0.88	0.88	0.88	88.06	μ-Avg.	0.86	0.86	0.86	85.95
	M-Avg.	0.90	0.87	0.88		M-Avg.	0.91	0.86	0.87		M-Avg.	0.90	0.84	0.86	
	W-Avg.	0.89	0.89	0.89		W-Avg.	0.89	0.88	0.88		W-Avg.	0.89	0.86	0.86	
RF	Neg	1.00	0.82	0.90		Neg	0.99	0.80	0.88		Neg	1.00	0.77	0.87	
	Neu	0.89	0.99	0.94		Neu	0.80	0.99	0.88		Neu	0.75	1.00	0.85	
	Pos	0.97	0.93	0.95		Pos	0.98	0.84	0.90		Pos	0.99	0.77	0.87	
	μ-Avg.	0.94	0.94	0.94	93.68	μ-Avg.	0.89	0.89	0.89	**89.00**	μ-Avg.	0.86	0.86	0.86	**86.73**
	M-Avg.	0.95	0.92	0.93		M-Avg.	0.92	0.87	0.89		M-Avg.	0.91	0.85	0.86	
	W-Avg.	0.94	0.94	0.94		W-Avg.	0.91	0.89	0.89		W-Avg.	0.89	0.86	0.86	
SVM	Neg	0.96	0.88	0.92		Neg	0.99	0.78	0.87		Neg	1.00	0.76	0.86	
	Neu	0.92	0.97	0.94		Neu	0.78	0.98	0.87		Neu	0.74	1.00	0.85	
	Pos	0.96	0.94	0.95		Pos	0.96	0.81	0.88		Pos	1.00	0.76	0.86	
	μ-Avg.	0.94	0.94	0.94	**94.21**	μ-Avg.	0.87	0.87	0.84	87.36	μ-Avg.	0.86	0.86	0.86	85.56
	M-Avg.	0.95	0.93	0.94		M-Avg.	0.91	0.86	0.87		M-Avg.	0.91	0.84	0.86	
	W-Avg.	0.96	0.94	0.94		W-Avg.	0.89	0.87	0.87		W-Avg.	0.89	0.86	0.86	

tations for text categorization. In *Proceedings of the 7^{th} ACM International Conference on Information and knowledge management*, pages 148–155, 1998.

[12] GuimarÃ£es, R., Rodríguez, D. Z., Rosa, R. L., and Bressan, G. Recommendation system using sentiment analysis considering the polarity of the adverb. In *2016 IEEE International Symposium on Consumer Electronics (ISCE)*, pages 71–72, 2016.

[13] GuimarÃ£es, R. G., Rosa, R. L., De Gaetano, D., Rodríguez, D. Z., and Bressan, G. Age groups classification in social network using deep learning. *IEEE Access*, 5:10805–10816, 2017.

[14] Havrlant, L. and Gil, J. M. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *International Journal of General Systems*, 46(1):1–21, 2017.

[15] Hofmann, H., Kafadar, K., and Wickham, H. Letter-value plots: Boxplots for large data. Technical report, had.co.nz, 2011.

[16] Junior, E. L. L., Rosa, R. L., and Rodríguez, D. Z. A recommendation system for shared-use mobility service. In *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6, 2018.

[17] Khurana, M., Gulati, A., and Singh, S. Sentiment analysis framework of twitter data using classification. In *Proceedings of IEEE 5th International Conference on Parallel, Distributed and Grid Computing, Solan, India*, pages 459–464, 2018.

[18] Kim, S. W. and Gil, J. M. Research paper classifcation systems based onÂ tf-idf andÂ lda schemes. *Human Centric Computation and Information Science*, 9(30):1–21, 2019.

[19] Kowsari, K., Jafari, K. M., Heidarysafa, M., and Brown, D. Text classification algorithms: A survey. *Information*, 10(150):6–10, 2019.

[20] Lasmar, E. L., de Paula, F. O., Rosa, R. L., AbrahÃ£o, J. I., and Rodríguez, D. Z. Rsrs: Ridesharing recommendation system based on social networks to improve the userâs qoe. *IEEE Transactions on Intelligent Transportation Systems*, 20(12):4728–4740, 2019.

[21] Liu, B. *Handbook of Natural Language Processing*. CRC Press, New York, second edition, 2010.

[22] M. Ghiassi, D., J. Skinner. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert System with Applications*, 30(16):6266–6282, 2014.

Figure 12: Representation of the best results achieved by the RF and SVM on different (a) uni-gram (b) bi-gram and (c) tri-gram models on Dataset-I

[23] Michalski, R. S. A theory and methodology of inductive learning. *Artificial Intelligence*, 20(2):111–161, 1983.

[24] Naz, S., Sharan, A., and Malik, N. Sentiment classification on twitter data using support vector machine. In *Proceedings of IEEE International Conference on Web Intelligence, Santiago*, pages 676–679, 2018.

[25] Pranckevicius, T. and Marcinkevicius, V. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, 10(150):6–10, 2017.

[26] Rani, S. and Gill, N. Hybrid model for twitter data sentiment analysis based on ensemble of dictionary based classifier and stacked machine learning classifiers-svm, knn and c5.0. *Journal of Theoretical and Applied Information Technology*, 98(4):624–634, 2020.

[27] Rosa, R. L., Rodríguez, D. Z., and Bressan, G. Sentimeter-br: A new social web analysis metric to discover consumers' sentiment. In *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, pages 153–154, 2013.

[28] Rosa, R. L., Rodríguez, D. Z., and Bressan, G. Sentimeter-br: A social web analysis tool to discover consumers' sentiment. In *2013 IEEE 14th International Conference on Mobile Data Management*, volume 2, pages 122–124, 2013.

```
Support Vector machine:

[[ 996  101   35]
 [  21 2343   52]
 [  25  113 2314]]
              precision    recall  f1-score   support

    Negative       0.96      0.88      0.92      1132
     Neutral       0.92      0.97      0.94      2416
    Positive       0.96      0.94      0.95      2452

   micro avg       0.94      0.94      0.94      6000
   macro avg       0.95      0.93      0.94      6000
weighted avg       0.94      0.94      0.94      6000

0.9421666666666667
```

Predicted Class

SVM	Negative	Neutral	Positive
Negative	996	101	35
Neutral	21	2343	52
Positive	25	113	2341

Actual Class

(a)

```
Random Forest:

[[ 900  207   25]
 [   2 2391   23]
 [  10  393 2049]]
              precision    recall  f1-score   support

    Negative       0.99      0.80      0.88      1132
     Neutral       0.80      0.99      0.88      2416
    Positive       0.98      0.84      0.90      2452

   micro avg       0.89      0.89      0.89      6000
   macro avg       0.92      0.87      0.89      6000
weighted avg       0.91      0.89      0.89      6000

0.89
```

Predicted Class

RF	Negative	Neutral	Positive
Negative	900	207	25
Neutral	2	2391	23
Positive	10	393	2049

Actual Class

(b)

```
Random Forest:

[[ 876  251    5]
 [   3 2404    9]
 [   0  567 1885]]
              precision    recall  f1-score   support

    Negative       1.00      0.77      0.87      1132
     Neutral       0.75      1.00      0.85      2416
    Positive       0.99      0.77      0.87      2452

   micro avg       0.86      0.86      0.86      6000
   macro avg       0.91      0.85      0.86      6000
weighted avg       0.89      0.86      0.86      6000

0.8608333333333333
```

Predicted Class

RF	Negative	Neutral	Positive
Negative	876	251	5
Neutral	3	2404	9
Positive	0	567	1885

Actual Class

(c)

Figure 13: Representation of the best results achieved by the RF and SVM on different (a) uni-gram (b) bi-gram and (c) tri-gram models on Dataset-II

[29] Rosa, R. L., Rodríguez, D. Z., and Bressan, G. Music recommendation system based on user's sentiments extracted from social networks. *IEEE Transactions on Consumer Electronics*, 61(3):359–367, 2015.

[30] Rosa, R. L., Rodríguez, D. Z., Schwartz, G. M., de Campos Ribeiro, I., and Bressan, G. Monitoring system for potential users with depression using sentiment analysis. In *2016 IEEE International Conference on Consumer Electronics (ICCE)*, pages 381–382, 2016.

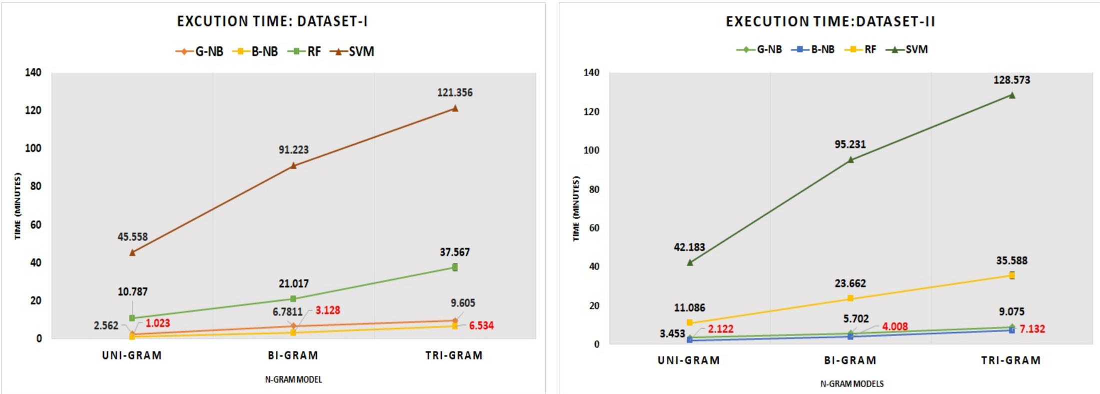
[31] Rosa, R. L., Schwartz, G. M., Ruggiero, W. V., and Rodríguez, D. Z. A knowledge-based recommendation system that includes sentiment analysis and deep learning. *IEEE Transactions on Industrial Informatics*, 15(4):2124–2135, 2019.

[32] S. Grover, S. A. Twitter data based prediction model for influenza epidemic. *In Proceedings of IEEE 2nd International Conference on Computing for Sustainable Global Development, India*, pages 873–879, 2015.

[33] Sailunaz, K. and Alhajj, R. Emotion and sentiment analysis from twitter text. *Journal of Computational Science*, 36(101003):1–18, 2018.

[34] Samuel, J., Ali, G. G., Rahman, M., Esawi, E., and Samuel, Y. Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, 22(4):1–21, 2020.

[35] Sarah, F. Forecasting zika incidence in the 2016 latin america outbreak combining traditional disease surveillance with search, social media, and



(a) (b)

Figure 14: Execution time taken by G-NB, B-NB, RF and SVM on different N-gram models on (a) Dataset-I (b) Dataset-II

news report data. *PLoS neglected tropical diseases*, pages 6–10, 2017.

[36] Silva, D., Rosa, R. L., and Rodríguez, D. Z. Sentimental analysis of soccer games messages from social networks using userâs profiles. *INFOCOMP*, 19(1), 2020.

[37] Singh, A. K. Applying the basic sir model to understand the outbreak of covid-19. *INFOCOMP Journal of Computer Science*, 19(1):1–5, 2020.

[38] Singh, R. and Singh, R. Machine learning based twitter data mining to analyse sentiments of tweets allied to covid-19 epidemic & its patterns. *ITEE, Information Technology & Electrical Engineering*, 9(2):11–22, 2020.

[39] Sokolova, M. and Lapalme, G. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 42(5):427–437, 2009.