SOCIAL MEDIA ANALYSIS WITH AI: SENTIMENT ANALYSIS TECHNIQUES FOR THE ANALYSIS OF TWITTER COVID-19 DATA

Rijwan Khan[1], Piyush Shrivastava[2], Aashna Kapoor[3], Aditi Tiwari[4], Abhyudaya Mittal[5]

[1]Professor & Head, Department of Computer Science and Engineering, ABES Institute of Technology, Ghaziabad, INDIA

[2]Scientist- 'C', National Institute of Electronics and Information Technology (NIELIT), Guwahati, INDIA

[3,4,5]Scholar, Department of CSE, ABES Institute of Technology, Ghaziabad, INDIA

ABSTRACT: Recently, In the current situation there has been an outbreak known as COVID-19 (corona virus) causing acute respiratory syndrome, first noticed in China and now a pandemic. Social media plays a crucial role in the current scenario of the world being locked up and further leading to the social imbalance among people. The news scattered like leaves about people attempting suicide.

In this chapter, we aim at providing the sentiment analysis on covid-19, about the people's reaction towards the decisions made either by the government or the local authorities through Twitter. We propose a system which automates analyzing the tweets and categorizing them into positive, negative or neutral sets. With the utilization of automata and NLP (natural language processing) together the accuracy, quantization and prediction of the sets can be achieved. Classification can be whether on a pattern based or a NLTK (Natural language toolkit). The classified results are further stored in the structures that could be iterated while calling for the visualization.

KEYWORDS: Social Media, Sentiment Analysis, Covid-19, Naive-Bayes classifier, Feature vector

I. INTRODUCTION

Social Media platforms give chance to the public to express their emotions. These applications work on real-time data, Twitter, Facebook, Instagram and many more are such examples through which people can vote their expressions publicly. In this Pandemic Situation of Covid-19, people round the globe are sharing their opinions and emotions through these platforms. Every country is giving its best to fight against corona virus but the situation is still difficult for everyone to handle. There's a trauma among people during this pandemic regarding the virus & step taken for its prevention. Social Media Sentiments means thoughts, attitude, and opinions towards any situation. We are focusing on hidden Sentiments behind the people's reaction because social media is the most easily accessed platform in this situation to express.

Micro-blogging sites connect people all over the world by lending a helping hand in exchanging the opinions, moments and views about the situation. It has been noticed that while communicating over social media somehow the way of thinking and the sentiments of the people are also exposed, which helps in categorizing the people according to their views.

Twitter being considered among one of the world's largest social networking platform. Using real-time data from Twitter, storing them into understandable format and further analysing these datasets using NLP (Natural Language Processing) and Machine Learning. By using Sentiment analysis (It's additionally alluded as subject analysis, opinion mining, and appraisal extraction.), authors will be detecting the text polarity which will be helping us to arrange assumptions into three categories which are Positive, Neutral and Negative. Instigated by the way of this, the study executed by us is to use sentiment evaluation to display the general mood of the people and hit upon the opinions and feelings of the people about this pandemic. Not all frameworks are able to identify opinion or sentiments from utilization of non-formal language and this could yearn the examination and dynamic procedure [1]. Human language is difficult to be understood by machines, thus it calls for the need of Natural Language Processing. It is a branch of artificial intelligence that helps to analyse, interpret, and evaluate human understandable data into machine processing data.

Using algorithms like SVM, Naive Bayes for data analytics. During this pandemic of corona virus. NLP and information extraction (IP) are the two approaches to which sentiment analysis can take advantage in order to analyse the sentiments of users. During this pandemic sentiment analysis could be a helping hand for predicting the mental state of user during this pandemic situation.

This process of sentiment analysis incorporates three different strategies.

- **Document level classification** – In this strategy we can categorize the whole document in positive, negative and neutral.
- **Sentence level classification** – In this strategy classification of each sentence takes place as positive, negative and impartial.
- **Feature level classification -**The strategy where document/sentence can be classified individually on the basis of particular aspect.

It's evaluated that about 80 percent of the world's data is unformed, as such it is disorganized. Gigantic number of text data is made on daily basis however it is very difficult to break down, interpret, also monotonous and exorbitant. So, sentiment analysis, be that as it may, assists organizations with understanding this unformed data via naturally labelling it.

Opinion mining utilizes different Natural Language Processing (NLP) strategies and calculations, which we'll go over in more detail in next segment. The main algorithms are as follows-

- **Rule Based –**systems that perform opinion mining dependent on a lot of manually created rules.
- **Automatic -** systems that are dependent on machine learning techniques.
- **Hybrid-** systems that join both standard based and programmed approaches.

Hence, by doing sentiment analysis on twitter we'll get to know the bits of knowledge of what individuals think about Covid-19 and its effects. Also, analysing user's tweets which will offer assistance in knowing the real opinions of users.

II. RELATED WORK

From past few periods, a ton of work has been finished in improvising opinion mining method, and lot more is in process. The use of algorithms of deep learning and also machine learning for prediction, modelling, training, and imitating human-like behaviour has increased. Sentiment analysis is also one such part where we analyse the data and predict the outcome. Data is the biggest asset today and almost all data related to a person from his/her identity to the way he/she thinks is available on social sites. This has led to massive research conducted by various companies and organizations in order to understand people and their needs more efficiently. Views by the users can be found flooding over the platforms regarding any situation and can be considered as sentiments. These sentiments of people are being studied and analysed by tech giants and scholars using various approaches such as Naive Bayes method, linear regression, and other deep learning algorithms. With the result obtained from this analysis, we can classify the reactions of people on the particular event further classifying in the form of positive, neutral or negative. This helps in greater understanding of the nerve of people and makes the system run accordingly.

A) Opinion Mining

Opinion mining alludes to the wide zone of NLP (that is natural language processing), content analysis that includes the computational think about of feelings communicated in content. it is used in order to recognize client assumption toward items, brands or administrations in online discussions and criticism.

According to the analysis of sentiments has numerous application spaces including bookkeeping, law, research, amusement, instruction, innovation, governmental issues, and showcasing [3]. In prior days numerous web-based life have given web clients road for opening up to communicate and impart their considerations and insights [4].

Understanding people's feelings is basic for businesses since clients are able to precise their contemplations and sentiments more straightforwardly than ever.

B) Twitter

A microblogging and social networking platform launched in 2006, through which users share and communicate through tweets.

Twitter, with almost 500 million clients and million messages for every day, has immediately turned into an important resource for associations to invigilate their notoriety and brands by removing and examining the estimation of the tweets by the general population about their items, administrations advertise and even about contenders [5].

Twitter could be a treasure trove of data. Numerous associations and clients tweet joins to curiously articles or unused data and the location can be an awesome device for youthful individuals to broaden their understanding of the world. Whereas nailing down particular reasons for Twitter's victory is troublesome, there are a few things which have contributed to it building up a enormous client base of over 500 million individuals all over the world.

Twitter has genuinely set up itself as a developing outlet for the spread of data. A portion of the top shortenings are Fb is used for Facebook whereas B4 is used for previously, OMG for oh my god, etc. In this way nostalgic investigation for the very short tweets like post of twitter is testing [17].

C) Social Media

Characterized an online networking as a gathering of Internet-put together applications that make with respect to the ideological and mechanical establishments of Web2.0 which is permitted to construct and trade of client produced substance. The control of web-based social networking is the ability to partner and offer information with anyone on Earth, or with various individuals simultaneously [6].

Online life may take the state of a grouping of empowered activities. These activities fuse photograph sharing, blogging, social gaming, social frameworks, video sharing, exchange frameworks, virtual universes, reviews and substantially more. For sure, governments and officials use online networking to secure with constituents and voters. For organizations, it is an indispensable instrument. Organizations use the phase to find and lock in with customers, drive bargains through publicizing and progression, and promoting customer advantage or backing.

A portion of the well-known web-based life stages are Twitter, Facebook, Instagram, YouTube and so forth.
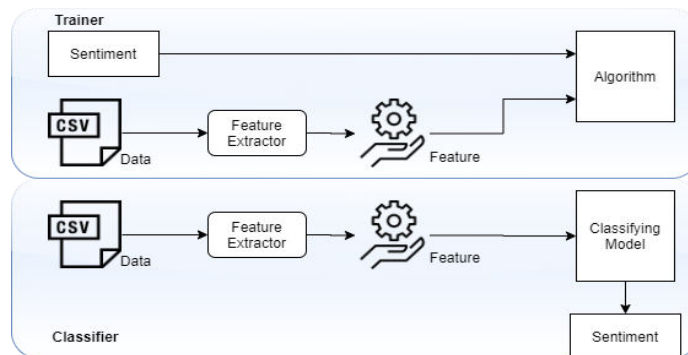
D) Twitter Sentiment Analysis

Sentiment analysis basically a technique of natural language processing (NLP) that helps in to quantifying an opinion that has been expressed in the tweet [7].Sentiment analysis is a general method with the help of which one can easily extract the polarity and subjectivity from the strength of words and phrases and text polarity [8].

We can automatically extract the sentiments with the help of the following approaches

- **Lexicon based Approach**
- **Machine learning based Approach**

In sentiment analysis using twitter we analyse different tweets and categorize them into positive, negative and neutral.

Fig 1:   Training & Classification

E) Sentiment Analysis Technique

For assigning the polarity we use certain techniques under sentiment analysis and these techniques are as follows.

1-Natural Language Processing (NLP)
The area of artificial intelligence research is Natural language processing (NLP) that fills in as a field of utilization and cooperation of various other conventional AI territories [9].

NLP is a field in software engineering which incorporates causing PCs to get significance from human tongue and contribution as a method of cooperating with the certifiable world.

NLP includes various areas like sentiment analysis, text classification, machine translation, speech recognition.

2-Case-Based Reasoning (CBR)
Case-based reasoning (CBR) is a straight thinking and experimental strategy dependent on the recovery and adjustment of cases, which are rambling depictions of issues and their related arrangements [10].

Case-based thinking has been formalized for reasons for PC thinking as a four-advance handle.

- **RETRIEVE** – Target problem is provided, retrieve the memory cases significant to solve it.

- **REUSE-**Mapping the target problem with the solution.
- **REVISE-** Testing the solution in the real world and revising it if necessary.
- **RETAIN-**After the target problem has adapted the solution successfully. we will be storing the result in memory as a new case.

3-Artificial Neural Network (ANN)

ANNs are considered nonlinear factual information modeling instruments where the complex connections between inputs and outputs are modeled or designs are found. The relation between the input data and the output data can be found using ANN [21, 22].

[11] Reference has been made to the possibility that the Artificial Neural Network (ANN) are viewed as the neural system a scientific procedure that interrelates the group of neurons which are artificial.

4-Support Vector Machine (SVM)

We can detect the sentiments of user's tweets through this support vector machine method. Support vector machines (SVMs) are effective however adaptable supervised machine learning algorithms which are utilized both for classification and regression. But by and large, they are utilized in classification issues. Every passage in the vector relates to the nearness a component.[18]

F) Python

A high-level, object oriented, interpreted and general-purpose programming language developed by Guido Van Rossum famously known as the creator of python programming language. This language provides increased productivity and more over programmers find this language very easy and interesting due to which it is one of the most loved languages.

 [12] referenced that Python is a language that is incredible for composing a model since Python is less tedious and working model gave, diverge from other programming dialects. When it comes to making a project, which is somewhat complex then programmers prefer this language. Some of the features of python programming language are as follows.

- a. SIMPLE
- b. EASY TO LEARN
- c. OBJECT ORIENTED
- d. INTERPRETED
- e. OPEN SOURCE
- **f. EXTENSIVE LIBRARIES**

Code reuse and modularity are the features which makes it more interesting as it supports modules and packages.

G) Application Programming Interface (api)

Application programming interface permits two applications to associate with one another. As the Python Twitter API just included Twitter messages for the latest six days, gathered the information should have been put away in an alternate database [13].

Python can naturally determine recurrence of messages being retweeted like clockwork, arranged the main 200 messages dependent on there-tweeting recurrence, and put away them in the assigned database [5]. Also, APIs are used by many programmers to build or create software.

API has a boundary that determines which language to recover tweets in. We generally set this boundary to English. Along these lines, our arrangement will just deal with tweets in English [2] since our preparation information is English-as it were.

H) Microblogging with E-Commerce

In later a long time, the boundaries between e-commerce and social organizing have gotten to be progressively obscured. Numerous e-commerce websites back the instrument of social login where clients can sign on the websites utilizing their social organize personalities such as their Facebook or Twitter accounts. Twitter has restricted for few words which are intended for the fast transmission of data or trade of sentiment [14].

[11] stated, actually, the organization's assembling such items have begun to survey propositions microblogs to get a feeling of general assessment for an item. In earlier times there was a wide boundary between the Microblogging and E-commerce but as we moved forward with time that boundary started becoming blurred.

We can see that many e-commerce has started using social logins where users can use their social network identities to sign in to the e-commerce websites. With the help of which the users nowadays are making purchases directly from the websites and all they have to do is to signup directly through social identities.

III.   SYSTEM OVERVIEW

The idea is to develop a system that outperforms state of the art method to recognize and classify the sentiments from the recorded datasets, considering the both understandable and non-understandable data. As the proposed framework is based on NLP algorithms which reduces the complexity for the system to process the data. For the feature extraction the polarity and subjectivity methods are used.
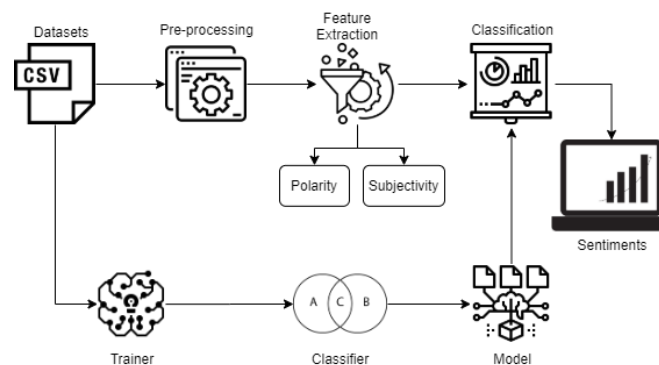

Fig 2:   Proposed System Strategy

IV. METHODOLOGY

Pre-Processing of data is termed as cleaning and preparing the text for classification and processing. Whenever a person tweets something on twitter, that tweet contain details such as hash tags, URLs, username, and a lot more. Our primary focus is to analyse the sentiments of people regarding COVID-19 pandemic. So, one of the main tasks is to modify the data according to the need. Thus, this modification becomes necessary as irrelevant information creates a hindrance in accurate analysis of data.

A)   Importing Data
The datasets for the initial processing can be gathered from the twitter developer account using their APIs. A sample of about 50,000 tweets tweeted by the people regarding the COVID-19 pandemic from February to April was analysed. The extracted data was stored in a csv or similar format that is the program's input. This data contains all the information tweeted by the user but only a specific set of information is required to perform a sentiment analysis. The Pandas library is used to import data into the data-frame from the CSV. Then date is updated to the required format i.e.; (month / day/year) by string slicing.

B)   NLTK Module
The most common algorithms are contained in NLTK and these algorithms are such as tokenizing, grammatical form labelling, stemming, assumption examination, point division, and named substance acknowledgment. The steps after the collection of data includes breaking of data into tokens or segments, removing the stopwords, and performing the stemming so that it can be easily processed. For this purpose, various functions from the Natural Language Toolkit (NLTK) Module is used. With the help of this module, the data of tweets can be divided into words. This provides access to individual word from the data and analyse the relevance of these words in our sentiment analysis process. The NLTK modules are discussed below.

i)   Tokenization
Tokenization is the process of breaking down the long paragraph of texts into smaller parts, such as words or sentence. Token is a single entity that for sentence or paragraph build blocks [19]. It is very convenient as the imported data has long sentences which make it difficult to analyse and process that data. So, to easily perform the various operations we break the data into tokens. Given below are the various tokenization methods.

a) Sentence Tokenization

Sentence Tokenizer helps in breaking the paragraph into corresponding sentences. It helps in converting the long paragraphs into smaller sentences.

Example: 'He is analysing the sentiments of people. He is working on sample data set.' - ['He is analysing the sentiments of people.', 'He is working on sample data set.']

b) Word Tokenization

Word tokenizer splits a paragraph of text into words. Hence all the words of a sentence can be separately stored and processed.

Example: 'He is working on a sample dataset' -

['He', 'is', 'working', 'on', 'a', 'sample', 'dataset']

ii) Stopwords

Texts that contains words like is, am, are, this, a, an, etc. Stopwords in the text are considered to be noise. So, it is always better to get rid of such stopwords because they do not contain any additional data. These words do not add any significance to the data as they are general words and they do not express any feeling or emotion of people. NLTK module has a library which contains list of all possible stopwords. So, all stopwords must be removed by comparing each word of our data with the words present in the existing library.

Example: ['He', 'is', 'working', 'on', 'a', 'sample', 'dataset'] - ['working', 'sample', 'dataset']

iii) Stemming

In the expectation that the 'base' form of a word will be found, Stemming removes the end of a word (-ing, -ed, -s, or another common end) .It is necessary because a lot of words are mostly used in different forms and tenses which usually have same meaning but spellings are different. This makes it more difficult and tiring to analyse or process the words. Hence, stemming is necessary before processing the data.

Example: ['connecting' - 'connect', 'helped' -'help']

C) Regex Module

The data collected for analysis contains a lot of irrelevant information, such as hash tags, usernames (@), hyperlinks, RT (symbol when a tweet is retweeted), emoticons (☺, ☹) etc. These symbols and words do not convey any meaning during the processing of data as they have nothing to do with the sentiment analysis. They just contain user information, tags to another person, or information related to the type of tweets. Therefore, raw twitter data must be structured in order to construct a dataset that can be easily accessed by different classifiers.[20]. The Regex Module is used to clean the data. Regex, or Regular Expressions, are special syntax strings that allow us to match patterns in other strings. There is a module in python called're' to work with regular expressions. It contains a number of methods for cleaning irrelevant data, it removes hash tags, URLs, retweet, hyperlinks, emoticons from data. The following sets of text and symbols are replaced using the regex:

i) Uniform Resource Locator (URL)

URLs and hyperlinks are removed because they do not contribute analysis of the tweet. All the mentioned URLs in the messages or tweets are replaced with blank space. To match the URLs we use regular expression and that is ((www\.[\S]+)|(https?://[\S]+))[20].The punctuation marks and digits / numerals are also removed as we want to compare the tweet to the English word list.

Example: ['https', ':', '/', 'dataset', 'sentiment'] -

[' ', ' ', ' ', 'dataset', 'sentiment']

ii) User Mention

Every user of the twitter has a handle assigned to them. Users mention other people in their tweets by using @handle. Hence all the user mentions are also removed. Talking about the regular expression (regex) that are used to match the mention of the user is @[\S] +. [20]

Example: '@name' - 'name'

iii) Hashtags

Hashtags are un-spaced sentences prefixed with the hash symbol (#) that users frequently use to refer a trend on twitter. All the hash tags are replaced with the words which contains hash symbol such as #correct is replaced by correct. The regular expression used to match hashtags is #(\S+). [20]

Example: '#covid-19' - 'covid-19'

iv) Lowercase

All data is converted in lowercase to make it standardized and homogeneous. Thus, making it easier to compare with the English dictionary.

Example: 'COVID-19' is converted into 'covid-19'.

v) Emoticons

Emoticons also known as Emojis have been created as a symbolic representation of emotions. To convey emotions the we see users using a wide number of different emoticons in their tweets. All emoticons are categorized in negative, positive and neutral. Thereby replacing all the emoticons with related words.[20]

The emojis contained in a tweet, their extremity, and whether the last badge of a tweet is an emoji are utilized highlights [19].

Example: 😊 used to express happy mood – replaced by specific word.

vi) Retweet

Many a times users share the tweets which are posted by other users and these shard tweets are known as retweets. The retweets start with the letters RT. We exclude RT from tweets because this is not an essential aspect of text classification. For retweet matching the regular expression which is used is \brt\b. [20]

Example: ['RT', '+', 'corona', 'virus', 'pandemic'] –

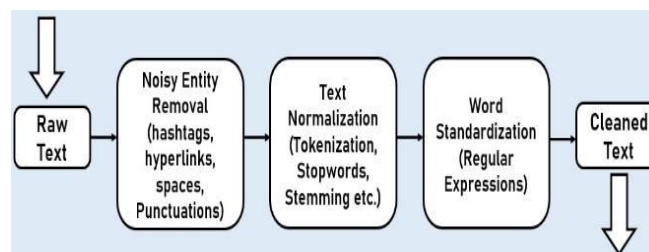 [' ', ' ', 'corona', 'virus', 'pandemic']

Fig 3:   Text cleaning Pipeline

V.   FEATURE EXTRACTION

Features are the various characteristics related to the data which can be used to distinguish it from other data. Feature Engineering is the extraction of these distinct features related to various sections of the data and group the similar features into similar categories. The extraction of features of the data under analysis is a very important step as features help in the study, analysis, and processing of the sentiments of people from the data. analysis of sentiments can be depicted as a programmed strategy, which incorporates mining of suppositions, perspectives, mentalities, and feelings from text, discourse, tweets through Natural Language Processing (NLP). The words assessments, Sentiment, convictions and perspectives are utilized interchangeably; however, contrasts exist between them [15].

- **Sentiment:** Opinion representing one 's thoughts.
- **Belief**: Intentional acceptance and intellectual concurrence.
- **View:** Subjective thinking.
- **Opinion:** A conclusion open to controversy
  (because different Experts think differently)

More than 100 features were analysed for this paper, whether they are positive, negative, neutral, facts, or opinions and that can be divided into two major categories, Polarity and Subjectivity. We used the Textblob library to analyse the polarity and subjectivity of tweets.

A) TEXTBLOB LIBRARY

Textblob is a python library which provides various functionalities to process textual data and perform basic Natural Language Processing (NLP) operations. It provides a simple API for diving into common NLP tasks such as part-of - speech tagging, word extraction, sentiment analysis, classification, translation, and more. The sentiment function of Textblob returns two properties, polarity as well as the subjectivity.so Text blob library is used to identify the polarity and subjectivity of the given data.

B) POLARITY

The main idea to perform sentiment analysis is to know how people feel, what they think about this situation and how they have expressed their emotions on the social platform. Hence, one of the main ways to analyse the features is to test whether they fall in the category of positive, neutral or negative. This is called the polarity of the data.

Textblob library is used to identify the polarity of the data by a simple representation:

Value	Polarity
-1	Extremely negative
0	Neutral
1	Extremely positive

In this way, the data is represented in the simple form by which the further analysis and processing can be done easily and the traits of data can be distinctly identified.

- **Polarity – Positive**

Example: This bacteria-killing UV light robot helps a Wuhan hospital diagnose #coronavirus patient. It disinfects its CT scanner to prevent cross-contamination

- **Polarity-Neutral**

Example: Head nurse Cai Liping of Wuchang Hospital died of #coronavirus infection in Wuhan. When the van transporting her body drove away, a man tried to grab the van and see her one last time. That's her husband, a doctor working for the same hospital.

- **Polarity-Negative**

Example: Lost one close family friend due to #COVID19.he was suffering from Kidney Ailments as well. Om Shanti. Be safe &At Home.

C) SUBJECTIVITY

People share a lot of information on the social sites to express their views and emotions. Apart from knowing whether they convey which type of polarity, it is also necessary to understand whether the information is a known fact (objective) or just an opinion (subjective) of that person. This is termed as the subjectivity of the data. It helps to differentiate whether the data is a fact or opinion, thus finding that people are influenced by which type of thoughts.

It provides a simple notation to represent the data through which it can be easily identified whether it is a fact or opinion. This representation is given below:

| Value	Subjectivity
0	Fact
1	Opinion

a) Fact: A fact is a statement that is true and can be objectively verified or proven to be true.
For Example: Common symptoms of corona virus are respiratory symptoms with fever, fatigue, cough and very rapid or laboured breathing.
b) Opinion: An opinion, is a statement that holds the dimension of belief, that tells how someone thinks. The belief is not necessarily true and cannot be proven.
For Example: Bathing in cow dung can cure corona virus.
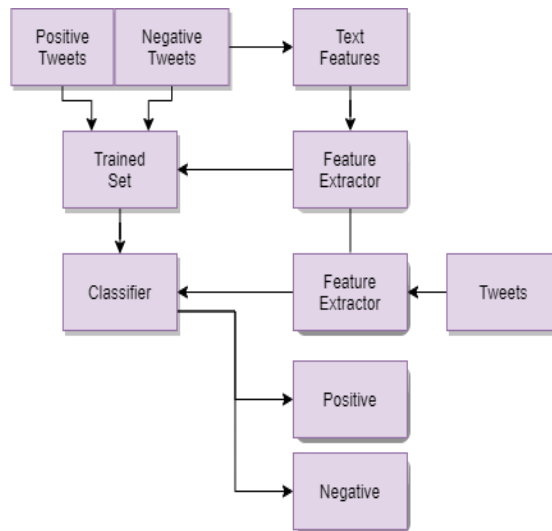
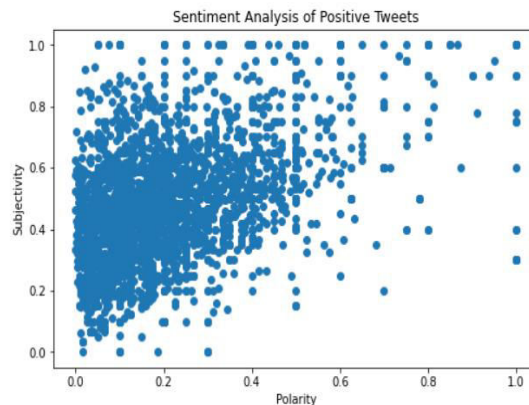**Fig 4:**   Feature extraction method for textual data

VI.   RESULTS & ANALYSIS

During the pre-processing of the data through the regex the following data frame was generated:
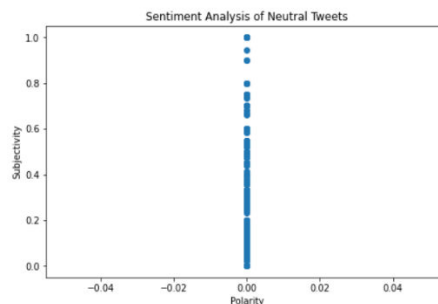

Fig 5: Pre-processed data-frame

A)   TextBlob Method

Using the TextBlob's polarity and subjectivity module, data has been classified through the scatter plots i.e.;

Fig 6:   Positive Tweets (Scatter plot) – Represent the distribution of tweets classified as positive using polarity & Subjectivity


Fig 7:   Neutral Tweets (Scatter plot) – Represent the distribution of tweets classified as neutral using polarity & Subjectivity
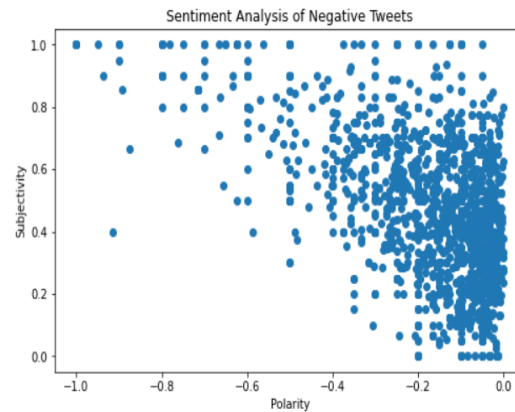
Fig 8: Negative Tweets (Scatter plot) – Represent the distribution of tweets classified as negative using polarity & Subjectivity

The bar representation below represents the total distribution of classified tweet percentage in the utilised datasets:
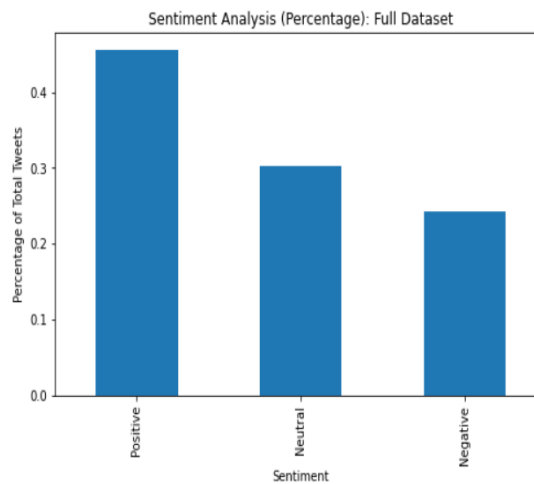
Fig 9:   Percentage of sentiments for the whole dataset

The monthly analysis for the data set describes the spread of sentiments over the time period where blue denotes positive, red denotes negative and grey represents the neutral sentiments:
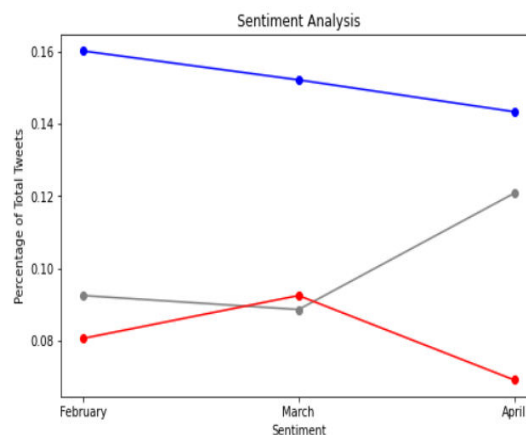
Fig 10:   Monthly sentiment analysis

**B)   Naive-Bayes Classifier**

This is a gathering procedure that relies upon Bayes' Theory with strong (honest) autonomy assumptions between the features [16]. Using the probabilistic technique for the accurate classification on the dataset, the trainer and classifier is deployed to process the data. The classifier on deployment generated the following results:

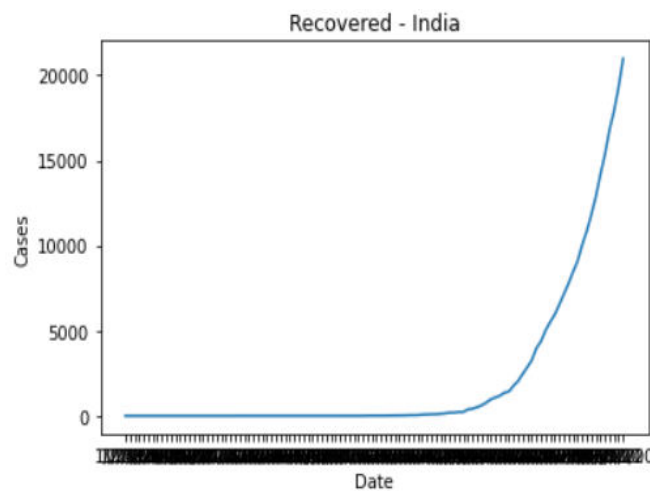Percentage of Positive Sentiments: 0.01953125

Percentage of Negative Sentiments: 0.0703125

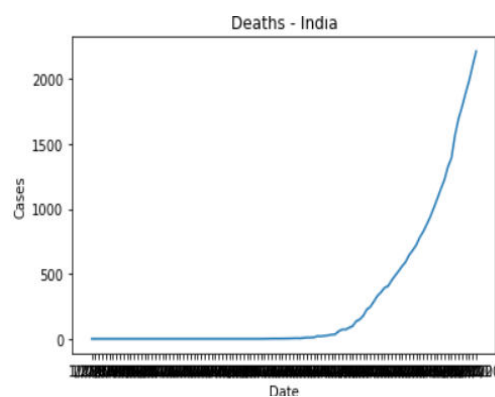**Fig 11:**   Percentage of classified tweets

We performed the analysis for the active user in the India regions and the results were as below:
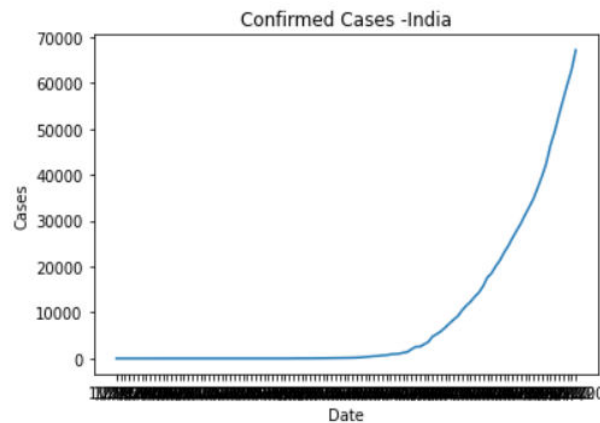
Fig 12: Confirmed cases in India- noted that the confirmed cases increased as the day passes
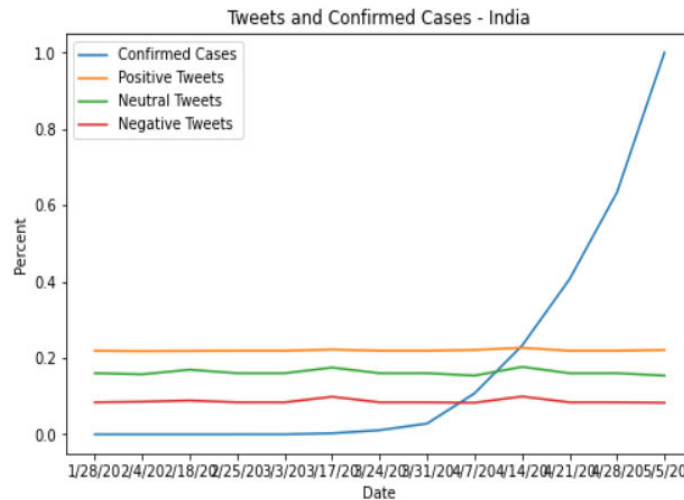

Fig 13:   Recovered cases in India- noted that the recovered cases increased as the day passes
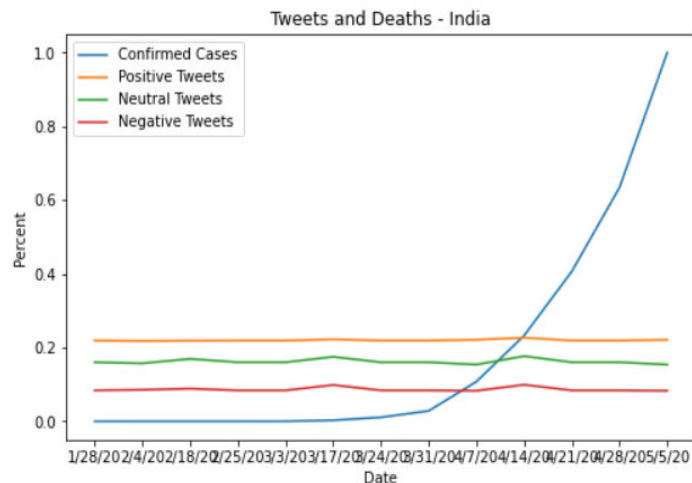

Fig 14:   Death cases in India- noted that the death cases increased as the day passes

It has been noticed that the tweets were shared constantly alongside the rising numbers of confirmed cases:

Fig 15:   Tweets shared Vs Confirmed cases

It has been noticed that the tweets were shared constantly alongside the rising numbers of death cases:

Fig 16:   Tweets shared Vs Death cases

A correlation was seen among the pattern of the positive, negative and neutral sentiments from the tweets done by the active users. The corresponding sentiment achieved a higher score when compared with the same sentiment class but in a different situation i.e.; with rising numbers of confirmed, recovered and death cases.

	Cases	Normalized	Positive Sentiment	Neutral Sentiment	Negative Sentiment
Cases	1.000000	1.000000	0.154972	-0.321203	-0.184889
Normalized	1.000000	1.000000	0.154972	-0.321203	-0.184889
Positive Sentiment	0.154972	0.154972	1.000000	0.571763	0.792186
Neutral Sentiment	-0.321203	-0.321203	0.571763	1.000000	0.883378
Negative Sentiment	-0.184889	-0.184889	0.792186	0.883378	1.000000

Fig 17:   Correlation during Confirmed cases

	Cases	Normalized	Positive Sentiment	Neutral Sentiment	Negative Sentiment
Cases	1.000000	1.000000	0.164397	-0.316887	-0.175757
Normalized	1.000000	1.000000	0.164397	-0.316887	-0.175757
Positive Sentiment	0.164397	0.164397	1.000000	0.571763	0.792186
Neutral Sentiment	-0.316887	-0.316887	0.571763	1.000000	0.883378
Negative Sentiment	-0.175757	-0.175757	0.792186	0.883378	1.000000

Fig 18:   Correlation during Death cases

	Cases	Normalized	Positive Sentiment	Neutral Sentiment	Negative Sentiment
Cases	1.000000	1.000000	0.102878	-0.334373	-0.233100
Normalized	1.000000	1.000000	0.102878	-0.334373	-0.233100
Positive Sentiment	0.102878	0.102878	1.000000	0.600743	0.760407
Neutral Sentiment	-0.334373	-0.334373	0.600743	1.000000	0.934841
Negative Sentiment	-0.233100	-0.233100	0.760407	0.934841	1.000000

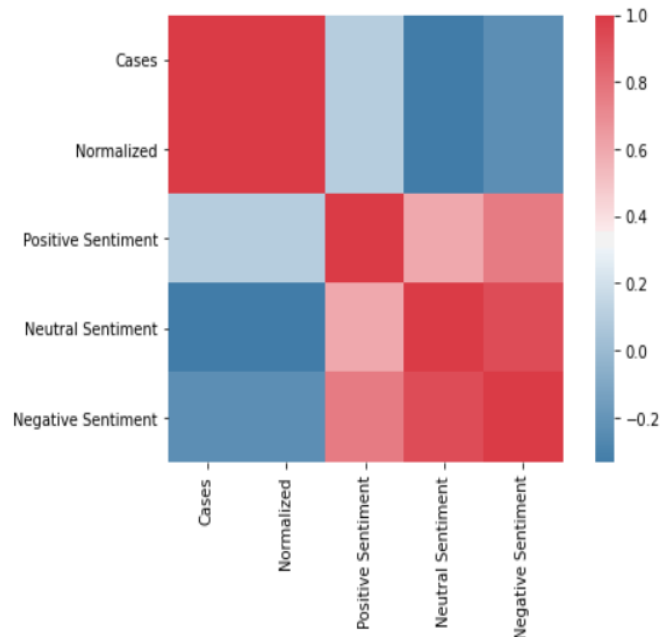**Fig 19:** Correlation during Recovered cases

Fig 20: Graphical representation for the Correlation

VII. RESULT and DISCUSSION:

While performing the analysis, it has been observed that by pre-processing the data using the regex and trainer has been an effective solution for clearing out the complexity of the applied algorithm as well the data rather than directly applying on the raw data itself. By using the trained model and further using it with the classifier proved to be a better way for classification as it reduced the time-frame and the size of the data-frame reducing the time complexity involved in the process. Talking about the analysis, it has been noticed that the number for tweets shared by the active users has been always greater as compared to the other sentiments. It means regarding the pandemic maximum number of people thought and took the decisions made by the government or the local authorities in a positive way. While the number of infected and demised people kept on increasing it didn't affect the mental strength of the population. For the 3-month analysis regarding the Indian sub-continent the variation among the positive, negative and neutral sentiments remained constant with the number of increasing cases day by day.

Currently the model is trained to study and analyse the sentiments of the people regarding Covid-19 pandemic. We can scale up this model to make it more user friendly so that rather than just being limited to covid-19, it may be further used to analyse the reaction/sentiments of people towardsthe major global events like ***#blacklivesmatter, #Brexit, #IndiaChinaFaceoff*** etc. Thus, for future scope/trends proposing a system which could store trained models for the dummy datasets of the global event. Users would be able to use the system by uploading their own datasets and filtering out the results by their own keywords related to the event, with the change of datasets the models would switch between themselves. This will also help in improving the algorithm behind the analysis and the system by adding new keywords as people will search according to their knowledge and by this way, shortcomings of the model can be found out and can be amended for further use.

VIII.   REFERENCES:

[1] Sarlan, C. Nadam and S. Basri, "Twitter sentiment analysis," Proceedings of the 6th International Conference on Information Technology and Multimedia, Putrajaya, 2014, pp. 212-216, doi: 10.1109/ICIMU.2014.7066632.

[2] A. Go, R. Bhayani, L.Huang. "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper ,2009

[3] D. Osimo, and F. Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis," Proceeding of the 12th conference of Fruct association, 2010, United Kingdom.

[4] A. Pak,and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," Special Issue of International Journal of Computer Application, France:Universitede Paris-Sud, 2010.

[5] H.Saif,Y.He,andH. Alani "SemanticSentimentAnalysisof Twitter," Proceeding of the Workshop on Information Extraction and Entity Analytics on Social Media Data. United Kingdom: Knowledge Media Institute, 2011.

[6] G. Kalia, "A Research Paper on Social Madia: An Innovative Educational Too", Vol.1, pp. 43-50, Chitkara University, 2013

[7] T. Carpenter, and T. Way, "Tracking Sentiment Analysis through Twitter,". ACM computer survey. Villanova:VillanovaUniversity, 2010.

[8] M.Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, " LexiconBased Methods for Sentiment Analysis," Association for Computational Linguistics, 2011.

[9] A. Gelbukh, "Natural language processing," Fifth International Conference on Hybrid Intelligent Systems (HIS'05), Rio de Janeiro, Brazil, 2005, pp. 1 pp.-, doi: 10.1109/ICHIS.2005.79.

[10] Allen, Bradley P. "Case-based reasoning: business applications." Communications of the ACM, vol. 37, no. 3, Mar. 1994, p. 40+. Gale Academic OneFile, Accessed 2 July 2020.

[11] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R.Passonneau, "Sentiment Analysis of Twitter Data," Annual International Conferences. New York:Columbia University, 2012.

[12] V.   Nareyko,   "Why   python   is   perfect   for   startups,"   Retrieved   01   10,   2014   from: http://opensource.com/business/13/12/why-python-perfect-startups

[13] J. Zhang, Y. Qu, J. Cody and Y. Wu, "A case study of Microblogging in the Enterprise: Use, Value, and Related Issues," Proceeding of the workshop on Web 2.0., 2010.

[14] D. Boyd, S. Golder, & G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter," System Sciences (HICSS), 2010 …. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=542831

[15] Kharde, Vishal & Sonawane, Sheetal. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications. 139. 5-15. 10.5120/ijca2016908625.

[16] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," Applied Intelligence, pp. 1-15, 2017.

[17] Trinh, S., Nguyen, L., Vo, M., & Do, P. (2016). Lexicon-based sentiment analysis of Facebook comments in Vietnamese language. In Recent developments in intelligent information and database systems (pp. 263-276). Springer International Publishing.

[18] A. Go, R. Bhayani, L.Huang. "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper ,2009

[19] M. Hagen, M. Potthast, M. Büchner, and B. Stein, "Webis: An ensemble for twitter sentiment detection," in Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), 2015, pp. 582-589.

[20] Twitter Sentiment Analysis System Shaunak Joshi Department of Information Technology Vishwakarma Institute of Technology Pune, Maharashtra, India Deepali Deshpande Department of Information Technology Vishwakarma Institute of Technology Pune, Maharashtra, India International Journal of Computer Applications (0975 – 8887) Volume 180 – No.47, June 2018.

[21] Yadav, S.P. and Yadav, S., 2020. Image fusion using hybrid methods in multimodality medical images. Medical & Biological Engineering & Computing, pp.1-19.

[22] Yadav, S.P. and Yadav, S., 2020. Fusion of Medical Images in Wavelet Domain: A Hybrid Implementation. Computer Modeling in Engineering & Sciences, 122(1), pp. 303-321.