Survey of Twitter Viewpoint on Application of Drugs by VADER Sentiment Analysis among Distinct Countries

Rajesh Bose[1], P. S. Aithal[2], & Sandip Roy[3]

[1] Post-Doctoral Research Fellow, Srinivas University, Mangalore 575 001, India
Orcid ID: 0000-0002-0967-455X; E-mail: bose.raj00028@gmail.com
[2] College of Management & Commerce, Srinivas University, Mangalore, India,
Orcid ID: 0000-0002-4691-8736; E-mail: psaithal@gmail.com
[3] Associate Professor, Brainware University, Kolkata, West Bengal 700125, India
Orcid ID: 0000-0002-5447-803X; E-mail: sandiproy86@gmail.com

March 2021

ABSTRACT

Who does not know that Twitter is an august social networking podium now? Here the folks around the globe are able to establish their viewpoints. Every day, almost 500 million tweets are established in twitter, and this volume contains 8TB data. The data that we derive from twitter is very much significant if we illustrate it, because we are able to derive significant news in the mode of sentiment analysis. From Twitter data, we get to know about the information and remarks of the augmentation of various product, novel fashion etc. Exploration of emotions, viewpoints, subjectivity and motive from a normal message or tweet for the application of drugs for the therapy of COVID-19, is the prime objective of this sentiment analysis. Now, comes clustering. It's a method by which one can detect homogeneous substances, combine together and create a class or cluster. There was an initiative from of going on with a research from our end and the result of it showed that the implementation of clustering is able to get infirm and solid positive or negative tweets while getting clustered with outcomes of distinct dictionaries and present robust scaffolding on our prediction. The research analyzes the polarity calculation, applying VADER sentiment analysis for the application of drug for the therapy of COVID-19.

Keywords: Dictionary based approach, Opinion Mining, Natural Language Toolkit, Sentiment Analysis, Social Media, Twitter, Valence Aware Dictionary and Sentiment Reason (VADER)

1. INTRODUCTION :

All are acquainted with this fact that the mechanisms of social networking sites remain in numerous, various ways like blogs, networks of trades, allotment of pictures, forums, micro blogs, business, social networks, allotment of video networks and social networks. Now, day by day, mechanisms of social network podiums have risen. Consequently, different social networking sites like Facebook, Twitter and YouTube are noteworthy now as these offer folks to present their viewpoints regarding life issues. If we look at US market, we can find that Facebook occupies 55% of it in 2020 and Pinterest gets the second place, occupying29.86 percent of it. In the countries of North America, specifically in USA, getting engaged in social networking is one of the noteworthy internet pursuits. If you calculate, you will find that 79 percents of internet users in USA use social networking profile. This assumed that the number of the members of social networking sites in USA will rise from 244 million in 2018 to over 257 million members in 2023. The graph as follows depicts the market scenario of social networking sites in United States in the span of May, 2020.
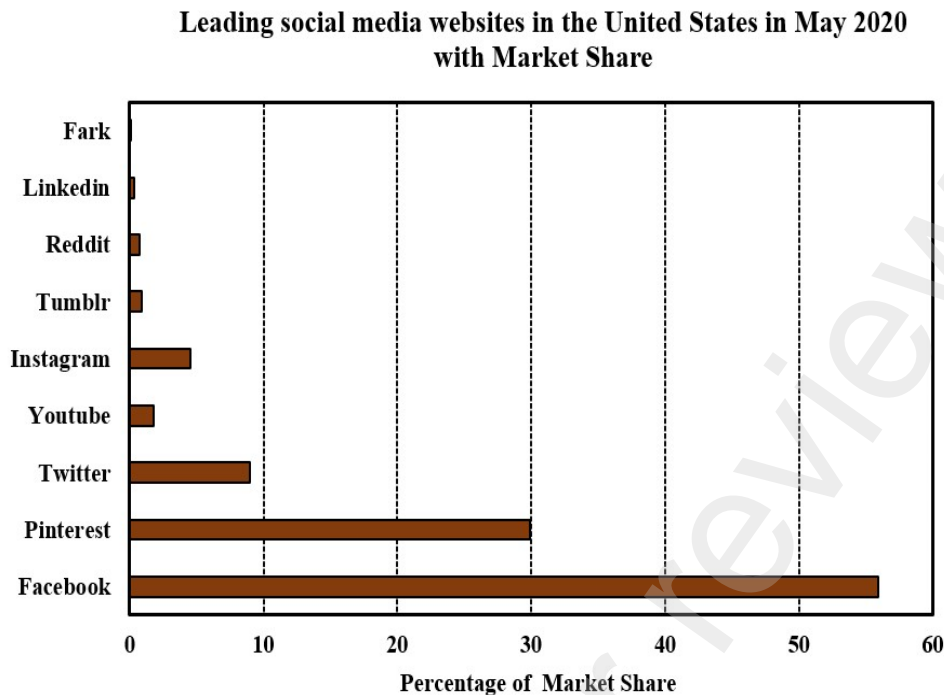
1

Leading social media websites in the United States in May 2020 with Market Share

Fig. 1: Market share of social media websites in USA

The networking sites allow their members to go on with conversations with divergent folks around the globe. They offer them as well to upload texts, photos and videos [1], [2]. Above all, social media have a huge area of providing business concerns with data for understanding common people's viewpoints and getting referendum about the products these concerns produce. Now, microblogging amenities are the most noteworthy and customarily applied podiums. Besides, these have developed to be imperative means of distinct details [3]. It is Twitter which is an august microblogging amenity which offers its members to allocate, convey and elucidate real-time in a brief and easy texts namely tweets [4]. Consequently, Twitter offers a liberal means of information which is applied in the ground of opinion mining and sentiment analysis. At present, the researches of this sphere are inclined to work with Twitter data. Nevertheless, the majority of thwarts have implemented sentiment analysis to derive and categorise data on the viewpoints presented in Twitter covering numerous events like assumptions, assessments, votes and business. At present several means like Linguistic Inquiry and Word count (LIWC) [5] allow the way to draw out novel attributes from messages. Nevertheless, majority of them demands a little programming skill. In this paper, we have implemented the Valence Aware Dictionary and Sentiment Reason (VADER) [6] for resolving the polarity of tweets and categorising those in terms of meticulous sentiment analysis. The rest of this research work is constructed below: Segment 2 presents an abridged illustration of analogous research regarding language. Segment 3 is an elaborate discussion on the suggested procedures. Here in this research, which device we have implemented, that is also discussed. Segment 4 is the illustration of the outcomes, and segment 5 is the conclusion part of this study and the guidance for prospective task.

2. BACKGROUND STUDY :

Divergent text extraction methods are implemented by derivation of data. The method of data processing of Twitter data was suggested by Dhirajgurkhe, Niraj Pal and Rishi Bhatia. They accumulated from various places. Then they removed the attribute that is not used to get any polarity. Next, these data are transmitted to the sentiment categorisation algorithm that computes the possibilities such as what amount of data is rectified and assumed the sentiment for specific questions [7]. The suggestion of Mr. Bouazizi, T. Ohtsuki is of the tweets having more than one sentiments and it is named as multiclass sentiment analysis. How they have detected the appropriate sentiment, is stated by the user. He

2

abstains from stating the entire sentiment of the tweet. For detection of it, they have applied SENTA device too. They suggested a means. They have computed the sentiment score, using this device. Thus, whose sentiment obtains highest score is called "Quantification" [8]. Geetikagautam, Divakar Yadav have applied tagged Twitter dataset for illustrating consumer assessment categorization. For doing such work, they have applied machine learning concerning algorithm such as Naive Bayes, SVM, maximum entropy. There performance was there on python and NLTK for workout of the SVM, Naive Bayes, and maximum entropy. If you concentrate on precision, Naive Bayes is certainly superior method and produces more effective outcomes in comparison with maximum entropy. Thus, more effective outcome can be produced by SVM we make comparison of it with unigram prototype. Next, another upcoming precision can be upgraded by semantic analytic pursued by WorldNet [9]. An exquisitely fitting prototype has been illustrated by Akshay Amolik, Niketanjivane, Mahavir Bhandari and Dr. M Venkatesan in their research work in which they will extract the Twitter data of forthcoming Hollywood and Bollywood films. They can do this work using clarifier and attributes such as SVM and Naive Bayes. Those are applied for super appropriateness. On the contrary, in case of appropriateness, Naive Bayes is superior to SVM. Now, if we discuss on recollection, SVM is superior to Naive Bayes. With the augmentation of dataset, we are able to augment the categorisation aptness too [10]. An integrated method was illustrated by Subhabrata Mukherjee et al., which is called Twisent that elucidates the conflict such as spam tweet, realistic, noisy message, and newscast. Twisent comprises in auditor of terms and administrator of realistic measure. Auditor of terms looks for the noisy message while realistic measure controls the reality in tweets. Twisent offers us more effective outcome in comparison with C-feel-IT approach. The precision of having negative sentiment of Twisentsyem is recorded as the C-feel-IT [11]. In this research, Dmitry Davidov, Oren Tsur, Ari Rapport suggested an overseen sentiment categorisation model that acts aptly in case of Twitter data. Their application was there of K closest neighbour and attribute vector. The fundamental objective of this model conveys the detection and differentiation between the kinds of sentiments that are illustrated by emoticons and tags [12]. Writer Neethu M.S Rajasree R has applied the machine learning devices in this scrutiny research for discovering the Twitter data that are in connection with electronic components. They have implemented attribute vector for the categorisation of tweets. Here they have applied three kinds of clarifier such as SVM, Naive Bayes, and maximum entropy. Those clarifiers trailed with the application of Matlab simulator. SVM, Naive Bayes clarifiers are used with the application of fabricated tasks, while MaxEnt clarifier is implemented by MaxEntsoftware. Consequently, each classifier produces almost the similar outcome [13]. The creation and suggestion of a prototype by Pulkit et al. is used to derive tweet from Twitter by reason of updates regarding terrorism. They went on with the research regarding terrorist intrusion that took place in URI on September 18 in 2016. They surveyed 59,988 tweets that had collected after the intrusion. Just those tweets are surveyed by them, that have #UriAttack, #uriattack, #uriattacks. In such cases, they applied Naive Bayes and SVM to draw out the latest Re-tweet time and counting of Re-tweets [14]. There is a suggestion of method of Sudarshan Sirrat et al. in sentiment analysis on Twitter data. There they have accumulated analysis of the brands. In such cases, they have implemented Naive Bayes algorithm that works better in question of precision and performance. 200 tweets have been drawn out in this event, whose average expanse is 70.105. The objective of that study is detection of the attributes of tweet such as the number of times it was liked and the number of times the tweet has been Re-tweeted [17]. Here comes Hetu etal's creation and suggestion of a prototype in sentiment analysis regarding Twitter data concerning anaconda python! In this event, the dataset from Kaggle has been drawn out, where they categorise the motives of folks concerning positive and negative analysis. This prototype offers a great precision on huge dataset [16]. Then we will discuss on the suggestion of a prototype offered by Ali Hasan et al. He applied the amalgamated method which consists of sentiment analyser machine learning. They had just collected such tweets which has hashtag (#) and comprise in latest views of politics. Fundamentally, this prototype translates into English tweets from Urdu tweets. In this event, 1690 tweets were collected for training data. 400 were for checking them. Here the Naive Bayes and SVM clarifiers have been implemented for training the dataset in weka and devising a prototype. In this occasion, distinct libraries have been applied to compute the subjectivity and polarity [15]. Now, we will talk about Feddah Alhumaid Al Otabi et al's suggested prototype. In this regard, he

3

applied the administered and unadministered algorithm. With the implementation of sentiment analysis, they expressed their wish to learn what food hub is superior in the eyes of the food lovers between McDonald and KFC. Besides, 700 tweets of those two food hubs were drawn out by Twitter API. The script and language used in this tweet was English and they have observed that MaxEnt has produced more effective outcomes in comparison with other means. Besides, it has been observed that in favour of KFC, several neutral tweets are there whereas McDonald has received more positive and negative tweets [14].

3. SENTIMENT ANALYSIS ON TWITTER :

Such method fundamentally comprises in four layers such as tweet retrieval, tweet-pre-processing, classification of algorithm and evaluation. In the following figure, the layers are briefly illustrated.
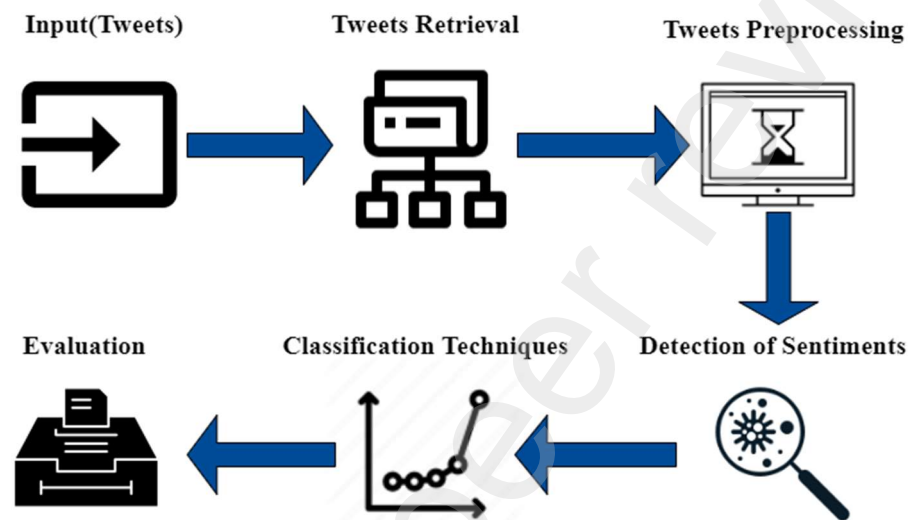

Fig. 2: Stages of Sentiment Analysis

3.1 Input (Keyword):
At first, we have to select the context. We will accumulate the tweets associated with that context. Next, on the basis of these tweets, we have to proceed with sentiment analysis.

3.2 Tweets Retrieval:
This layer states that tweets will be got back. The structure of it can be either configured or semi con-figured or jumbled manner. In sentiment analysis, experiment is executed with the application of distinct programming language such as python or R. Thus we are able to accumulate the tweets.

3.3 Pre-processing:
This layer offers the pre-processing of the filtration of data, eliminating the unnecessary data, conflict-ing data and noisy data. In the time of pre-processing, the works mentioned below are accomplished:
- Eradication of re-tweets (twitter data sets)
- Eradication of specific attributes, URLs, numbers and punctuations etc.
- Eradication of ceasing words
- Derivation
- Tokenization

3.4 Identification of Sentiments: Sentiment analysis offers different kinds of uses in which this is compulsory to get the sentiment such as tweet categorization and tweet derivation. In this event, we search the polarity of the provided word. It is either positive or negative or neutral. Applying distinct

4

glossaries, here we are able to detect the polarity such as Bing Lui sentiment, glossaries, sentiWord Net and some more. They will assess sentiment score, its stability and many more [18].

3.5 Sentiment classification algorithm:

If we make a categorisation of sentiment analysis then we shall get two proposals. They are administered learning and unsupervised learning. Administered learning demands Naive Bayes, SVM and maximum entropy to carry out the sentiment analysis while unsupervised learning demands the learning on the basis of glossary aggregation and dictionary to carry out the sentiment analysis. Besides clarifier's precision is used on the basis of the nature of training and checking dataset implemented.

3.6 Assessment:

In the event of sentiment analysis, concluding output is illustrated for coming to this point if it ought to be chosen. After that, the outcome is shown depicting graph, pie chart and line graph.

4. LEXICON BASED APPROACH UNSUPERVISED SENTIMENT CLASSIFICATION TECHNIQUES :

In this research, the experimental method comprises in the accumulation of Twitter corpus and the categorisation with the application of glossary. When categorisation is done, we derive the amount of positive, negative and neutral opinions from the glossary. This is used to study the users. Use of glossary is a very significant method to sentiment analysis. This made use of computing the sentiment from the semantic point of reference of words or phrases which are found in a write-up or message or opinion [19]. A dictionary of positive or negative words is needed along with it, with a positive or negative sentiment worth that is allotted with each word. The suggestion of distinctive methods for the formation of dictionaries has been given, such as manual [20] and automatic [21] methods. Customarily, in the methods on the basis of glossary, a single text message is constituted by profuse words. On the basis of the constitution of the text, sentiment worth from the dictionary is allotted to each positive and negative word or phrase in the text. An amalgamated task like sum or average is used for getting the concluding notion about the entire sentiment for the text. Besides, the facet of the local subject-matter of a word is customarily regarded as denial or development. In this research, it is resolved that a method on the basis of glossary will be used for ignoring the requirement to produce a tagged training set. The chief demerit of machine learning prototype is their dependence on tagged data. Giving confirmation of acquiring adequate and aptly tagged data is almost impossible. Apart from it, there is no doubt of it that a method on the basis of glossary can be comprehended more smoothly and improved by a man. It is regarded an important merit for our research. It is observed that it is simpler to produce a suitable glossary than to accumulate and tag compatible corpus. It is stated that the date drawn out from social media are generated by users around the world. It would be a restriction if the algorithm is solely able to manage English language. That's why, there ought to be a simpler method of conversion of various languages, with the use of sentiment analysis algorithm. Afterwards, in the research work, it is illustrated how a sentiment analysis algorithm on the basis of glossary, can be fitted to various languages by a suitable translation of the sentiment glossary and use of string resemblance work. The method on basis of glossary has three categories. They are manual, dictionary based and corpus-based methods. Manual based takes much time to perform. Besides, it is restricted to a few glossaries and is likely to make error. For creating it well organized, it is amalgamated with other two mechanized methods.

4.1 Dictionary based approach:

Polarity signifies a series of opinions that is accumulated manually. After that it is developed well with the application of noteworthy copora WorldNet [22] or thesaurus [23] that extract the meaning of words along with their synonyms and antonyms. After that he lately extracted words are connected with the record and the iteration repeats itself. The procedure goes on until no novel words are derived. Taboada applied the method on the basis of glossary for deriving sentiments from micro blogs in the period of managing denial and growing words [24]. Methods on the basis of dictionary are unsupervised by character. Customarily, these means speculates that positive or negative adjectives emerge very often

close to a positive (negative) seed word (Harb et al., 2008) [25]. T categorise reassessments, an unsupervised learning algorithm (thumbs up or thumbs down) has been taken on by (Turney, 2002) [26]; Wangand Araki, 2007) [27]. A reassessment categorisation is specified by the average semantic direction of their phrases that comprise in either adjectives or adverbs. A phrase semantic direction is calculated with the application of the mutual information between the specified phrase and the word exceptional deficiency of the joint information between the specified phrase and the word poor.

$$SO(phrase) = \log \frac{hits(phrase\_NEAR\_excellent) \circ hits(poor)}{hits(phrase\_NEAR\_poor) \circ hits(excellent)} \qquad (1)$$

The following flowchart shows how the words applied to manifest opinions are comprehended. For this, positive and negative conceptual words such as superior, inferior, rich, poor etc. are applied to draw out adjectives which are close to conceptual words. The flowchart as follows, is the layers ofsentiment6 analysis on the basis of dictionary as discussed.
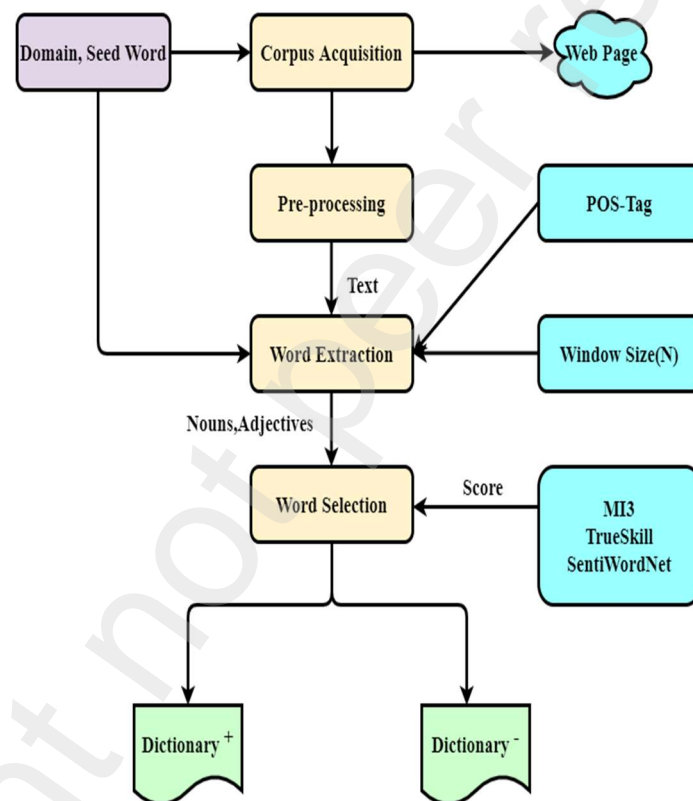

Fig. 3: Stages of dictionary-based sentiment analysis approach

4.2 Corpus based method:

It is a means that is on the basis of aggregation, is characteristically formed for promoting area particularised semantic glossary from a compilation of area particularised content. If newfangled dictionary method is used for exclusive vocabularies, it shows a few important disputes based on logic. That's why, Rice and Zone [28] modified a category of "minimally-supervised" with the intention of the formation of sentiment dictionary from aggregated write-up. In one use, its functionality is displayed in US federal appellate court settlement.

6

(1) Corpus-Based Statistics-Oriented (CBSO) Approach:

Customarily, this method has the attributes as follows:

1. This offers effective form that is used in customary languages in place of surface strings, for replicating the hypothetical attitude of the linguistics for the intention of diminishing the number of specifications in the linguistics prototype.
2. This applies a specialized analytical method to determine obscurities with the intention of the operation of objective development and necessary knowledge such as the specifications worth is got spontaneously and frequently.
3. This is customarily more solid in comparison with the absolutely analytical methods as analytical development is used on effective form of which analytical features are more appropriate to be concluded to invisible data better than surface settings.

The CBSO methods are blended methods. They take support of law-dependent and absolute analytical methods. A CBSO method creates analytical linguistic prototypes at the peak of effective forms like interpreted system trees. As a reference, we can say, a CBSO prototype for machine translation was suggested in [3][31][32] on the basis of effective forms like parse trees and general constructions, the translation dispute is replicated as a development dispute that opts the finest translation which augments the translation score below:

(1) $P(T_i \mid S_i)$

$$\cong \sum_{Ii} \{[P(T_i \mid PT_t(i)) \times P(PT_t(i) \mid NF1_t(i)) \times P(NF1_t(i) \mid NF2_t(i))] \quad (2)$$

(2) $\times [P(NF2_t(i) \mid NF2_s(i))]$

(3) $\times [P(NF2_s(i) \mid NF1_s(i)) \times [P(NF1_s(i) \mid PT_s(i)) \times P(PT_s(i) \mid S_i)]\}$

Here (S_i, T_i) is the i-th foundation-goal translation pair, (PT_s, PT_t) are the parse trees for the foundation-goal sentences, $NF1_s$ and $NF2_s$ signifies linguistically and semantically regularized parse trees, known as usual structures, of the foundation sentence, $NF1_t$ and $NF2_t$ are the usual structures of the aimed sentence, and the outline of the possibilities is seized all these kinds of transitional illustrations, I_i. The three equations (1), (2) and (3) by themselves delineate the generation, transmission and scrutiny prototypes of a tMT procedure on the basis of transmission in a CBSO approach; they are able to be again clarified for execution.(A few particulars are conveyed in the subsequent segments). The figure as follows, the steps of Corpus-Based Statistics-Oriented (CBSO) method is illustrated.
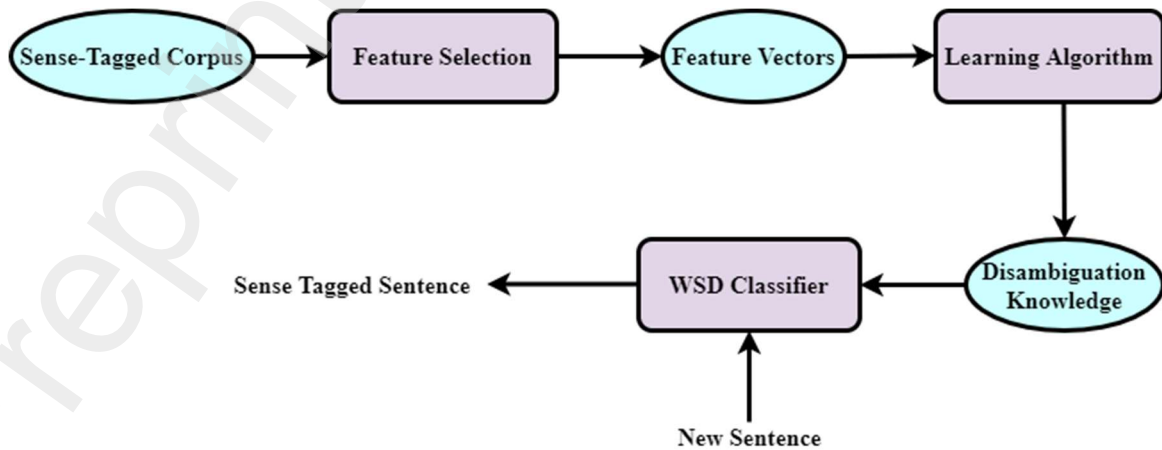

Fig. 4: Stages of Corpus-Based Statistics-Oriented (CBSO) approach

(2) Semantical Approach:

Corpus-based semantic prototypes approve the universal proposal that analogous words will be applied in analogous subject-matters and interpret the significance of a word or idea as a vector (ordered list) of worth which recapitulates, in a few means, the mark of the subject-matter of the word, that is, the series of subject-matters where the word turns up. In this research we have a few present some preparatory outcomes. The finest usual assumption is that if we delineate some distance metric on our semantic vector space. Next the vectors equivalent to semantically unattached words ought to be additionally spaced out than those for similar words. As there are several standard commodities which will perhaps come up in the time of making comparison of vectors mined applying dissimilar significance of the above considerations such as distinctive volume of standard commotion, the usual aspect liberated magnitude for making comparison is the similarity ratio:

$$R = \frac{\text{Mean distance between control words}}{\text{Mean distance between related words}} \tag{3}$$

If the ratio is bigger, then the similar words are nearer and the semantic exemplifications will be more appropriate. In this study, 100 pairs of words were taken. They were evaluated by subjects of human beings to be close synonyms [33]. Here for every single pair eight random pairs of words of similar frequency were taken to proceed with our management. After that, the co-occurrence vectors are formed and computed the ratio R applying easy Euclidean distances for such pairs for an assortment of consideration sets. The following diagram shows the layers of Semantical method.
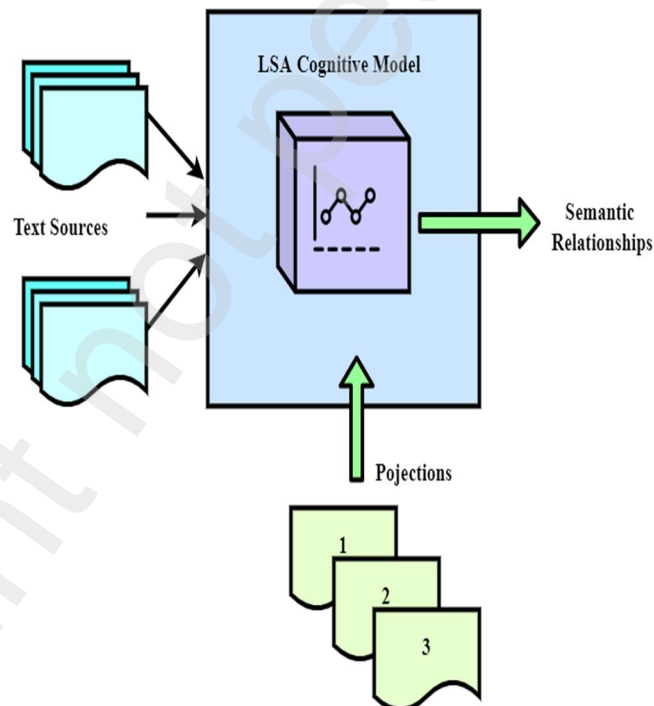

Fig. 5: Stages of Semantically approach

5. SUGGESTED PROTOTYPE ELEMENT ILLUSTRATIONL :

5.1 Data sets:

Twitter is nothing but a social networking podium which offers its members to transmit and interpret micro-blogs of up to 280-characters called "tweets". This helps registered members interpret and place their tweets via the web, short message service (SMS), and mobile features. The world real-time statistics communication podium declares that every month, more than 400 million members visit Twitter

8

and there are 255 million active members all over the globe. Among them numerous statesmen, august sport persons, prominent actors, business magnets etc. are there. At present you can access Twitter in more than 35 languages. Twitter set out its journey in 2006 by Jack Dorsey, Evan Williams, Biz Stone, and Noah Glass. Its headquarters are there in San Francisco, California, USA, with local offices in Atlanta, Austin, Boston, Boulder, Chicago, Detroit, Los Angeles, New York, Sunnyvale, and Washington. Its international offices are situated in Amsterdam, Berlin, Dublin, London, Madrid, Paris, Rio de Janeiro, Sao Paulo, Singapore, Sydney, Seoul, Tokyo, Toronto, and Vancouver.

The following tabular representation shows a model of the derived data in which on date and tweeted messages are displayed for our evaluation from the outcomes.

Table 1: Sample Tweet comments with post date and time

Date	Tweet Text
20-04-2020 04:39	#COVID-19: Lockdown can't Stop Virus -Prof Adesugba
20-04-2020 04:39	Tamil Nadu truly needed something like #OndrinaivomVaa to uplift its spirit and urge its people to fight against the COVID-19 pandemic as one!
20-04-2020 04:39	What can #Indonesia learn from the experience in the Republic of #Korea, one of the first countries to hold national #elections under the threat of the #COVID19 pandemic? Join the discussion on 21 April at 13.30-15.30 WIB:
20-04-2020 04:39	Lol. U are dressed in a mask and full gown in the streets screaming covid 19 is a lie??
20-04-2020 04:39	In the face of COVID 19 we want you all to take note on this and continue to stay safe. Let's stand together and save many lives as we can. #takeprecaution #staysafe #stayhome #savelives #helpstopthevirustoday
20-04-2020 04:39	On the Move: Strength, resolve, true leadership mark city's COVID-19 response READ MORE: #business #CEO #Change #Coaching #Communication #Conference #Development #Growth #Guidance #Influence
20-04-2020 04:39	#COVID?19 #coronavirus #lessonlearned #pandemic What have I learned from the COVID 19 crisis? 1. You're faster in shit than you think 2. There are more virologists in the world than I thought 3. You should grill the bat well before you eat it ??
20-04-2020 04:39	Dr. Birx said this a couple of weeks ago.
20-04-2020 04:39	Seat at home and Slay at home, COVID 19 said so, No more Hotels snapping Lolz....??
20-04-2020 04:39	My mom got Multiple Sclerosis 48 hours after her Hepatitis B vaccine. And she was one of many. I am not an "anti-vaxxer", I do believe in vaccines. But I have no trust in compulsory mass vaccination of a brand new vaccine. No COVID-19 vaccination for me. At least for a while. RT @SkyNews: Coronavirus: Novak Djokovic reveals he's an anti-vaxxer and it may stop his return to tennis
20-04-2020 04:39	Count them all to steer away from #COVID19 thoughts for a while! #nature #positivethinking
20-04-2020 04:39	The #SwissTech start-up Adamant Innotech and CSEM are developing a rapid, simultaneous test to determine if a person has been previously infected with #COVID19, influenza or SARS. #SwissTechL39
20-04-2020 04:39	Transferred from FB. Now will ppl believe that deaths from anything are recorded as covid 19 ?
20-04-2020 04:39	Prevention works!! We must make sure that prevention activities are not stopped during COVID-19 responses. RT @TheUnion_TBLH: A Union study in #Zimbabwe, published in @BMJ_Open shows reduction of #TB case notifications with the scale-up of antiretroviral and isoniazid preventive therapies. #EndTB #AdvancingPrevention #HIV

20-04-2020 04:39	#SouthAfrica postpone upcoming #SriLanka tour due to COVID-19 outbreak #COVID19 #Coronavirus
20-04-2020 04:39	The communities devastated by bushfires are still struggling, with people still waiting for the $2b promised to help them rebuild. @ScottMorrisonMP, we know COVID-19 is challenging, but you must not forget the bushfire victims.
20-04-2020 04:39	Interested in comparing global #COVID19 responses? "Coronavirus & liberty in Illiberal Europe" by @CamGeopolitics explores this in detail. Chaired by @Peterhouse_Cam Master @Bridget_Kendall & speakers @Heather-Grabbe Prof John Dunn & @Judy_Dempsey Link?? RT @dollytheis: Looking forward to this @CamGeopolitics digi event on Wednesday, 18:00-19:00?? "Coronavirus & Liberty in Illiberal Europe" Speakers: @Bridget_Kendall @Judy_Dempsey @HeatherGrabbe & emeritus Professor John Dunn Register here to join:
20-04-2020 04:39	How is everyone today? Things I'm not going to do today - Involve myself with Twitter-COVID drama - Read every single article ever written on COVID-19 - Write honest lists
20-04-2020 04:39	The so-called Gujarat Model. Is this how we will fight #COVID19? #PMWakeUpSaveIndia

5.2 Tools: Natural Language Toolkit (NLTK):

NLTK is nothing but liberated open-source software of the combination of Python which offers profuse devices to create programs and categorizes data. NLTK is appropriate software for various professionals such as linguists, engineers, students, educators, researchers, and developers who perform their duties with textual information in normal language processing and text analytics [34]. NLTK offers a simple means for the application of the interfaces of over 50 corpora and lexical foundations. This takes in a collection of text processing records for categorization, tokenization, and stemming, labeling, parsing, and semantic interpretations [35].

5.3 Valence Aware Dictionary and Sentiment Reasoner (VADER):

VADER is a glossary and law-dependent sentiment analysis tool which is exclusively standardized to the sentiments manifested in social media. This is a completely liberated open-source tool. VADER incorporates deliberation of word arrangement and degree qualifiers [36][37].

5.4 Pre-processing:

Each example of the data sets had phonetics and punctuation marks eliminated. Distinct pre-processing methods were used for the machine learning clarifiers.

- ✓ Lemma derivation: Each verb is shortened to the infinitive structure and nouns and adjectives to the singular male form. In this research we applied the device VADER (Valence Aware Dictionary and Sentiment Reasoner -a lexicon and rule-based sentiment analysis tool) [38].
- ✓ Stemming: It shortens all words to its fundamental. The Python documentation NLTK was applied to do so.
- ✓ Part of Speech (PoS) Labeling: the NLPNET Python documentation [39][40], that allows a PoS tagger for constituted texts, was applied.
- ✓ Summarization: Categorization stage demands solely the first three statements and the heading of the report. It is reliable with extractive briefing representations used in Web.

5.5 Algorithm and Flowchart of our Sentiment Analysis Model:

This segment shows the illustration of sentiment analysis prototype displayed in the flowchart as follows:
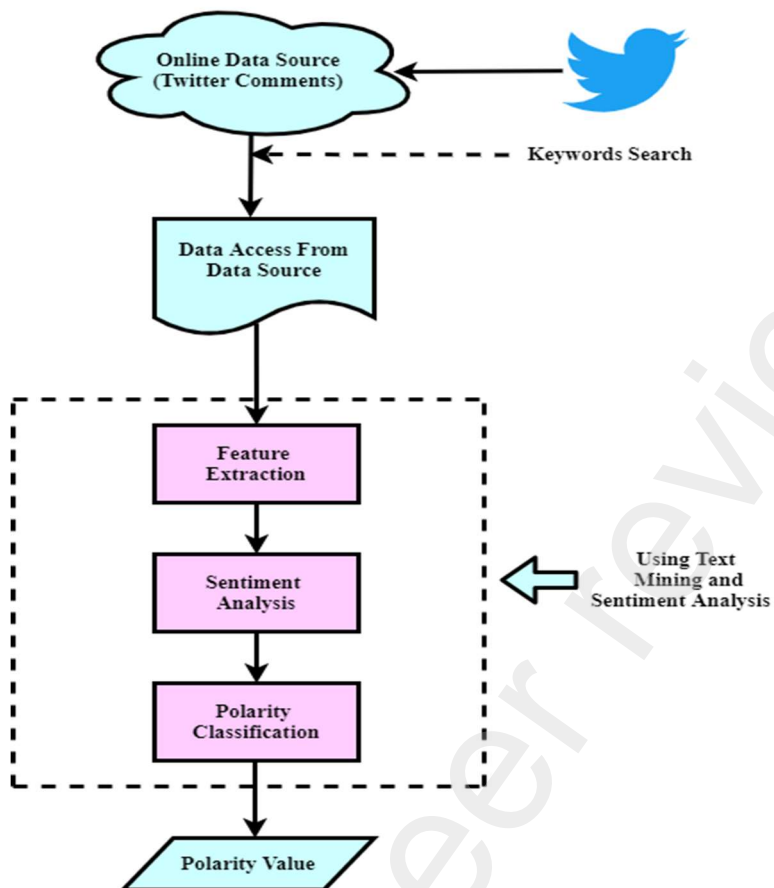
Fig. 6: Algorithm Flowchart

In the above figure stages to go on with the evaluation method are shown on the news articles so that we can classify the data about its positivity or negativity or spontaneity.

ALGORITHM 1: Twitter Comments Sentiment Classification using VADER

Input: Text File (Twitter Comments which include Nouns, Adjectives, Adverbs)

Output: Values > 0 (Positive), Values < 0 (Negative), Values = 0 (Neutral)

Begin:

1. Sentiment Analysis () ← File
2. **For** each row in rows
3.   **if** *Sentiment Polarity Score(line) > = 0.05* **then**
4.   Sentiment ← Positive
5.   **else**
6.   **if** *Sentiment Polarity Score (line) < = - 0.05* **then**
7.   Sentiment ← Negative
8.   **else** Sentiment ← Neutral
9.
10.   **end**
11.   **end**
12. **end**
13. **end**

5.6 Hydroxychloroquine and COVID-19:

World Health Organization has alarmed doctors for giving approval of dubious therapy to the people affected by COVID-19. WHO also has alarmed the patients affected by COVID-19, for taking self-chosen drugs? The specialists from all over the globe agrees on the3 point that the prospects subsist but vigorous research is required to come to this point whether subsisting antiviral medicines are fruitful for the treatment of the patients affected by COVID-19. If it gets fruitful then they could mitigate the hazards of COVID-19 from this world. In many countries, it is observed that COVID-19 has prevailed uncontrollably and has created ample anguish and mortality. When this kind of unusual state has arisen, physicians have given much effort to provide the patients affected by COVID-19, with every sort of treatment for trying to make improvement from this ailment. In the whole world, a particular therapy has been famous now, and that is, the factual application of chloroquine and hydroxychloroquine which are used for the treatment of malaria. Though it is not a proven fact yet whether the application of them is able to cure a patient affected by COVID-19, many countries have so far incorporated CQ/HCQ in COVID-19 treatment conventions [41], [42] for the means of a treatment choice for profoundly affected patients and for a preventive measure as well. Here a wide difference in the result of the study has been displayed for such all-inclusive analysis of the antiviral upshot of CQ/HCQ on SARS-COV-2 and some other viruses. It is observed that CQ and HCQ revealed favourable outcomes. Though some clinical efforts have brought a few prospective results of CQ/HCQ in treatment of the persons affected by COVID-19, the majority of the recorded are prefatory till date [43], [44], [45]. Besides, minimum 7 of the efforts in progress were invalid or ceased and this is not still evident whether it occurred on account of potential undesirable events, uselessness or any different issues. A number of cases regarding noxious elements is connected with such medicines [46], [47], [48], [49]. Among them, the leading alarming one is the probability of OT continuation and the danger of ventricular tachycardia, and it is a menacing arrhythmia for life [50], [51], [52], [53], [54]. However, the analysis of our paper establishes that it is not customary, and still this is not clear if any collaborative hazard is witnessed after amalgamating other medicines with such medicines like AZT [46]. Evidently, it is conflicting to go on with a settlement of therapy devoid of an absolute experimentation rhythm and a vivid idea of the success of medicines and security. As there is some unpredictability, we should keep in mind about the probable hazards. So, we should be firmly sensible about the application of such medicines to the persons affected by COVID-19 if super arbitrary medical efforts are not customary for illustrating the part of the therapy of recovery from COVID-19 [46].

6. RESULT ANALYSIS :

For managing evaluation from the outcome, we go ahead with calculation on the basis of data accumulated from Twitter data. For detection of positive or negative sentiments, we process article data. The following is such a tabular representation which the polarity along with the sentiment score as well as percentage is illustrated. These are accumulated from the dataset with the help of VADER sentiment analysis.

- **Twitter Sentiment Score Description of India:** In this tabular form the polarity and their sentiment score and percentage are shown. They are taken out by VEDAR sentiment analysis. Sentence Overall Rated is as "Positive".

Table 2: Polarity Value and Percentage of Polarity of Twitter Data Set of India

Polarity	Sentiment Score	Percentage
Positive	0.128	12.8
Neutral	0.776	77.6
Negative	0.096	9.6

- **Twitter Sentiment Score Description of USA:** In the following tabular form, the polarity with their sentiment score and percentage are depicted. Sentence Overall Rated is as "Positive".

12

Table 3: Polarity Value and Percentage of Polarity of Twitter Data Set of USA

Polarity	Sentiment Score	Percentage
Positive	0.133	13.3
Neutral	0.77	77.0
Negative	0.097	9.7

- **Twitter Sentiment Score Description of China:** In the following tabular representation, the polarity with their sentiment score and percentage are shown. Sentence Overall Rated is as "Positive".

Table 4: Polarity Value and Percentage of Polarity of Twitter Data Set of China

Polarity	Sentiment Score	Percentage
Positive	0.135	13.5
Neutral	0.756	75.6
Negative	0.109	10.9

- **Twitter Sentiment Score Description of Spain:** In the following tabular representation, the polarity with their sentiment score and percentage are shown Sentence Overall Rated is as "Positive".

Table 5: Polarity Value and Percentage of Polarity of Twitter Data Set of Spain

Polarity	Sentiment Score	Percentage
Positive	0.134	13.4
Neutral	0.766	76.6
Negative	0.10	10.0

- **Twitter Sentiment Score Description of Germany:** In the tabular representation as follows, the polarity with their sentiment score and percentage are illustrated. Sentence Overall Rated is as "Positive".

Table 6: Polarity Value and Percentage of Polarity of Twitter Data Set of Germany

Polarity	Sentiment Score	Percentage
Positive	0.138	13.8
Neutral	0.755	75.5
Negative	0.108	10.8

- **Twitter Sentiment Score Description of Italy:** In the tabular representation below, the polarity with their sentiment score and percentage are shown. Sentence Overall Rated is as "Positive".

Table 7: Polarity Value and Percentage of Polarity of Twitter Data Set of Italy

Polarity	*Sentiment* score	Percentage
Positive	0.136	13.6
Neutral	0.764	76.4
Negative	0.10	10.0

- **Twitter Sentiment Score Description of France:** In the following tabular representation, the polarity with their sentiment score and percentage are illustrated. Sentence Overall Rated is as "Positive".

Table 8: Polarity Value and Percentage of Polarity of Twitter Data Set of France

Polarity	Sentiment Score	Percentage
Positive	0.134	13.4
Neutral	0.769	76.9
Negative	0.098	9.8

- **Twitter Sentiment Score Description of African countries:** In the following tabular form, the polarity with their sentiment score and percentage are shown. Sentence Overall Rated is as "Positive".

Table 9: Polarity Value and Percentage of Polarity of Twitter Data Set of African Countries

Polarity	Sentiment Score	Percentage
Positive	0.183	18.3
Neutral	0.73	73.0
Negative	0.086	8.6

In the figure 7, the graphs depicted below, tell about the juxtaposition of distinct statesmen of different countries, offering an outcome, and also depict the mode of HCQs application concerning the therapy of COVID-19. Here we can come to this decision that the outcome of polarity assessment and subjectivity has a positivity aspect for the application of HCQs for the therapy of COVID-19. More experiments are yet to be made on the opinions or sentiments with the application of well-developed methodology.
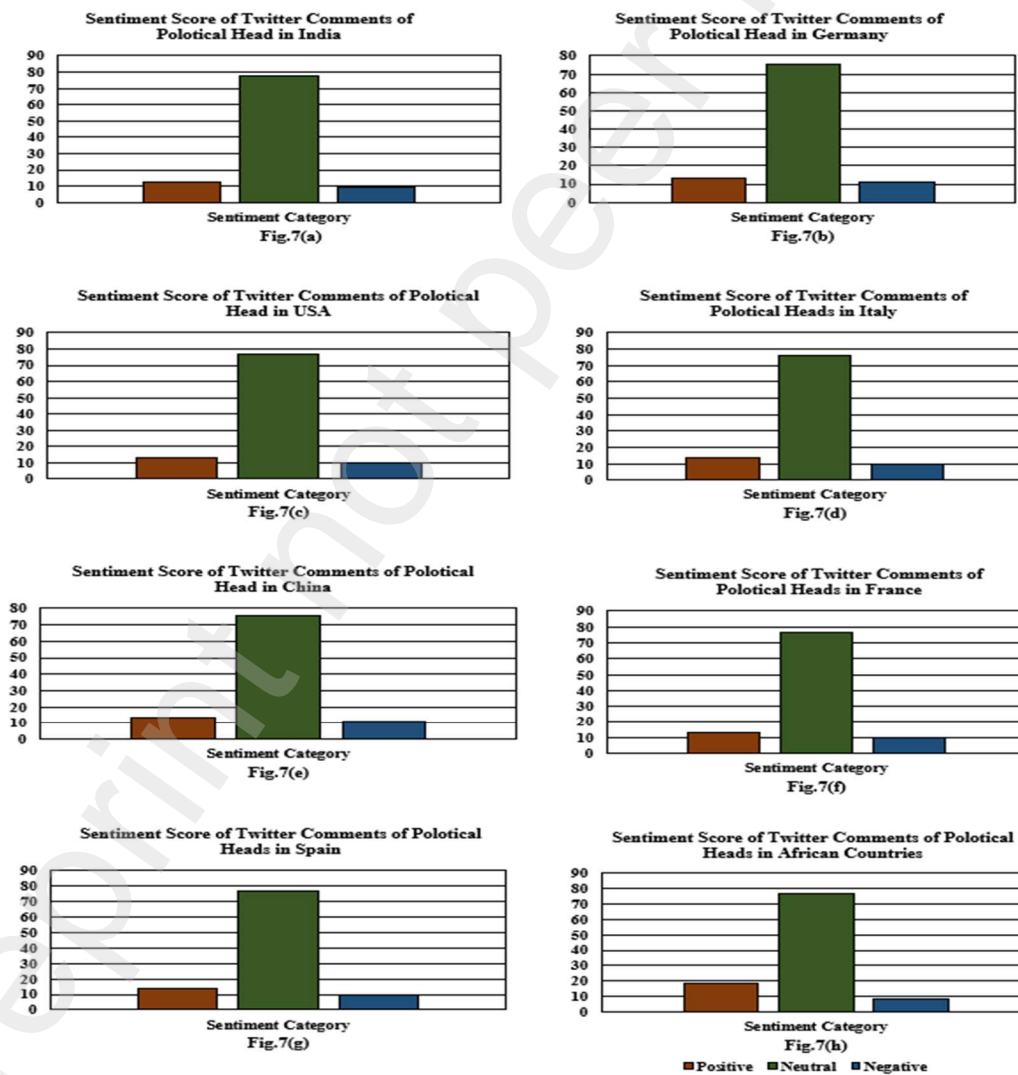

Fig. 7: Sentiment analysis result Twitter comments of different countries

7. CONCLUSION :

In this research wok, it is shown how the concurrent upshot of scrutinizing distinct countries' statesmen for comprehending the application of drug pattern with regard to COVID-19 is witnessed. The prototype has acted on the opinions of the tweets of the outcomes, this relative research is executed. This research work shows how VADER sentiment analysis approaches to make the prospective work. These outcomes display a concrete association of Twitter opinions with up/down in sentiment polarity. Additional experimentation can also be executed chain wise to have better outcomes in this area. In this paper, it will be examined too, what outcomes we will have, applying live instances of the consequence created by corona virus infections all over the globe. Profuse incidents with numerical approaches to comprehend the patterns common in COVID-19 circumstances, will be scrutinized on the basis of historical aspects. In this suggested prototype, views and opinions are posted by reviews on tweets will be assessed.

REFERENCES :

[1] Jansen, B. J., Zhang, M. Sobel, K. & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth, *J. Am. Soc. Inf. Sci. Technol.*, *60*(11), 2169 – 2188.

[2] Kharde, V. & Sonawane, P. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques, *Int. J. of Comp. Appl.*, *139*(11), 5 – 15.

[3] Selvaperumal, P. & Suruliandi, D. A. (2014). A short message classification algorithm for tweet classification, *2014 Int. Conf. on Recent Trends in Info. Tech.*, Chennai, 1 - 3.

[4] Singh, T. & Kumari, M. (2016). Role of Text Pre-processing in Twitter Sentiment Analysis, *Procedia Comput. Sci.*, *89*(1), 549 – 554.

[5] Tausczik, Y. R. & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods, *J. Lang. Soc. Psychol.*, *29*(1), 24 – 54.

[6] Hutto, C.J. & Gilbert, E. (2014). Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, *in Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, 1 – 10.

[7] Gurkhe, D., Pal, N. & Bhatia, R. (2014). Effective Sentiment Analysis of Social Media Datasets using Naïve Bayesian Classification. *Int. J. of Comp. App.*, *99*(13), 1 – 4.

[8] Bouazizi, M. & Ohtsuki, T. (2018). Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer. *IEEE Access*. *6*(1), 64486 - 64502.

[9] Gautam, G. & Yadav, D. (2014). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. *2014 Seventh International Conference on Contemporary Computing (IC3)*, 437 - 442.

[10] Amolik, A., Jivane, N., Bhandari, M. & Venkatesan, M. (2016). Twitter sentiment analysis of movie reviews using machine learning techniques. *International Journal of Engineering and Technology*, *7*(6). 1-7.

[11] Mukherjee S., Malu A., Balamurali A. R. & Bhattacharyya P. (2013). TwiSent: A Multistage System for Analyzing Sentiment in Twitter. *Proc. of the 21st ACM international conference on Information and knowledge management*, 1 – 4.

[12] Davidov D., Tsur O. & Rappoport A. (2010). Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. *International Conference on Computational Linguistics*, 241 – 249.

[13] Neethu, M. S. & Rajasree, R. (2013). Sentiment analysis in twitter using machine learning techniques. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1 – 5.

[14] Garg, P., Garg, H. & Ranga, V. (2017). Sentiment analysis of the Uri terror attack using Twitter. *2017 International Conference on Computing, Communication and Automation (ICCCA2017)*, 17 – 20.

[15] Hasan, A., Moin, S., Karim, A. & Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Math. Comput. Appl.*, *23*(1), 1 – 15.

[16] Bhavsar, H. & Manglani, R. (2019). Sentiment Analysis of Twitter Data using Python. *International Research Journal of Engineering and Technology (IRJET)*, *6*(3), 510 – 511.

[17] Sirsat, S., Rao, S. & Wukkadada, B. (2019). Sentiment Analysis on Twitter Data for product evaluation. *IOSR Journal of Engineering (IOSRJEN)*, *5th Somaiya International Conference on Technology and Information Management -SICTIM-2019*, 22 - 25.

[18] Behdenna, S., Barigou, F. & Belalem G. (2018). Document Level Sentiment Analysis: A survey. *EAI Endorsed Transactions on Context-aware Systems and Applications*, *18*(13), 1 – 8.

[19] Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Association for Computational Linguistics*, *37*(2), 267 – 307.

[20] Tong R. (2001). An operational system for detecting and tracking opinions in on-line discussions. In: Working Notes of the SIGIR Workshop on Operational Text Classification, 1 – 6.

[21] Turney P. D. & Littman M. L. (2003) Measuring praise and criticism: inference of semantic orientation from association. *ACM Transact. Inform. Syst.*, *21*(4), 315–346.

[22] Dey, S., Sarkar, I., Chakraborty, S. & Roy, S. (2020). DEPRESSION DETECTION USING INTELLIGENT ALGORITHMS FROM SOCIAL MEDIA CONTEXT - STATE OF THE ART, TRENDS AND FUTURE ROADMAP. J. of Xidian University, *14*(8), 45 – 62.

[23] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235-244.

[24] Mohammad, S., Dunne, C. & Dorr, B. (2009). Generating High-Coverage Semantic Orientation Lexicons from Overly Marked Words and a Thesaurus. *In: Proceedings of the 2009 conference on Empirical Methods in Natural Language Processing (EMNLP'09),* 599 – 608.

[25] Harb, A., Plantie, M., Dray, G., Roche, M., Trousset, F. & Poncelet, F. (2008). Web opinion mining: How to extract opinions from blogs? *In Proceedings of the 5th International conference on Soft computing as transdisciplinary science and technology (CSTST 08)*, 211–217.

[26] Peter, D. T. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL'02)*, 417 – 424.

[27] Guangwei, W. & Kenji, A. (2007). Modifying SO-PMI for Japanese weblog opinion mining by using a balancing factor and detecting neutral expressions. *In Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers, NAACL-Short'07*, 189 – 192.

[28] Rice, D. R. & Zorn, C. (2021). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods*, *9*(1), 20 - 35.

[29] Su, K.Y., Chiang, T.H. & Chang, J.S. (1996). An Overview of Corpus-Based Statistics Oriented (CBSO) Techniques for Natural Language Processing. *International Journal of Computational Linguistics and Chinese Language Processing*, *1*(1), 101–157.

[30] Su, K.-Y. & Chang, J.-S. (1990). Some Key Issues in Designing MT Systems. *Machine Translation*, *5*(4), 265 - 300.

[31] Su, K.-Y. & Chang, J.-S. (1992). Why Corpus-Based Statistics-Oriented Machine Translation. *Proceedings of 4th Int. Conf. on Theoretical and Methodological Issue in Machine Translation*, 249 - 262.

[32] Su, K.-Y, Chang, J.-S. & Una Hsu, Y.-L. (1995). A Corpus-based Two-Way Design for Parameterized MT Systems: Rationale, Architecture and Training Issues. *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95)*, 334 - 353.

16

[33] Moss, H. E., Ostrin, R. K., Tyler, L. K. & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*(4), 863 - 883.

[34] Lobur, M., Romanyuk, A. & Romanyshyn, M. (2011). Using NLTK for educational and scientific purposes. *2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, 426 - 428.

[35] Bird, S., Klein, E. Loper, E. (2009). Natural Language Processing with Python. *O'Reilly Media Inc.*, 1 – 438.

[36] Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2018). Sentiment Analysis on Online Product Reviews. *In Information and Communication Technology for Sustainable Development. Advances in Intelligent Systems and Computing*, *933*(1), 559 – 569.

[37] Valdez, D., Ten Thij, M., Bathina, K., Rutter, L. A., & Bollen, J. (2020). Social Media Insights into US Mental Health during the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data. *Journal of medical Internet Research*, *22*(12), e21418.

[38] Bose, R., Dey, R. K., Chakraborty, S., Roy, S., & Sarddar, D. (2019). Examining Hidden Meaning of E-commerce Platform. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(12), 257 – 261.

[39] Fonseca, E. R. & Rosa, J.L.G. (2013). A Two-Step Convolutional Neural Network Approach for Semantic Role Labeling. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, 1 - 7.

[40] Fonseca, E. R. & Rosa, J. L. G. (2013). Mac-Morpho Revisited: Towards Robust Part-of-Speech Tagging. *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, 98 - 107.

[41] Lenzer, J. (2020). Covid-19: US gives emergency approval to hydroxychloroquine de-spites lack of evidence. *BMJ*, *369*(1), 1 – 2.

[42] Galvis, V., Spinelli, F.-R., Tello, A., Sossa, C. L., Higuera, J. D., Gomez, E. D., Serrano, S. E., Camacho, P. A., & Velez, F. G. (2020). Hydroxychloroquine as Prophylaxis for Coronavirus SARS-CoV-2 Infection: Review of the Ongoing Clinical Trails. *Arch Bronconeumol. 56*(9), 606 – 608.

[43] Gao, J., Tian, Z., & Yang, X. (2020). Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Biosci Trends*, *14*(1), 72–73.

[44] Gautret, P., Lagier, J. C., Parola, P., Hoang, V. T., Meddeb, L., Mailhe, M., Doudier, B., Courjon, J., Giordanengo, V., Vieira, V. E., Tissot Dupont, H., Honoré, S., Colson, P., Chabrière, E., La Scola, B., Rolain, J. M., Brouqui, P., & Raoult, D. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int. J Antimicrob Agents, *56*(1), 105949.

[45] Chen, Z., Hu, J., Zhang, Z., Jiang, S., Han, S., Yan, D., Zhuang, R., Hu, B., & Zhang, Z. (2020). Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial". *medRxiv*, 1 - 11.

[46] Munster, T., Gibbs, J. P., Shen, D., Baethge, B. A., Botstein, G. R., Caldwell, J., Dietz, F., Ettlinger, R., Golden, H. E., Lindsley, H., McLaughlin, G. E., Moreland, L. W., Roberts, W.N., Rooney, T. W., Rothschild, B., Sack, M., Sebba, A. I., Weisman, M., Welch, K. E., Yocum, D., & Furst, D. E. (2002). Hydroxychloroquine concentration-response relationships in patients with rheumatoid arthritis. *Arthritis Rheum. 46*(6), 1460-9.

[47] Makin, A. J., Wendon, J., Fitt, S., Portmann, B. C., & Williams, R. (1994). Fulminant hepatic failure secondary to hydroxychloroquine. *Gut.*, *35*(4), 569–70.

17

[48] Murphy M., Carmichael, & A. J. (2001). Fatal toxic epidermal necrolysis associated with hydroxychloroquine". *Clin Exp Dermatol., 26*(5), 457 – 8.

[49] Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2019). Analyzing Political Sentiment using Twitter Data. *Information and Communication Technology for Intelligent Systems*. *Smart Innovation, Systems and Technologies*, *107*(1), 427 - 436.

[50] Stas, P., Faes, D., Noyens, P. (2007). Conduction disorder and QT prolongation secondary to long-term treatment with chloroquine. *Int. J. Cardiol.*, *127*(2), e80–2.

[51] Chorin, E, Dai, M., Shulman, E., Wadhwani, L., Bar Cohen R., Barbhaiya, C., Aizer, A., Holmes, D., Bernstein, S., Spinelli, M., Park, D. S., Chinitz, L. A., & Jankelson, L. (2020). The QT interval in patients with SARS-CoV-2 infection treated with hydroxychloroquine/ azithromycin. *medRxiv*, 1 – 11.

[52] Sardar, D., Dey, R. K., Bose, R., Roy, S. (2020). Topic Modeling as a Tool to Gauge Political Sentiments from Twitter Feeds. *IJNCR*, *9*(2), 1 – 22.

[53] Bose, R., Aithal, P. S., Roy, S. (2020). Sentiment Analysis on the Basis of Tweeter Comments of Application of Drugs by Customary Language Toolkit and TextBlob Opinions of Distinct Countries. *IJETER*, *8*(7), 3684 – 3696.

[54] Biswas, S., Ghosh, A., Chakraborty, S., Roy, S., Bose, R. (2020). Scope of Sentiment Analysis on News Articles Regarding Stock Market and GDP in Struggling Economic Condition. *IJETER*, *8*(7), 3596 - 3609.

****