BBRC
Bioscience Biotechnology
Research Communications

Customized Data Extraction and Effective Text Data Preprocessing Technique for Hydroxychloroquin Related Twitter Data

K.Priya[1]* and A. Anbarasi[2]
[1]*Department of Computer Science, AVP College of Arts and Science, Tirupur-64652, Tamil Nadu, India*
[2]*LRG Government Arts and Science College, Tirupur-641604, Tamil Nadu, India*

ABSTRACT

Coronavirus Disease 2019 (COVID-19) has become a significant worldwide issue with a rising the number of infected people and increases in mortality. Among every single helpful methodologies, contentions have raised about hydroxychloroquine (HCQ) viability in the treatment of COVID-19. The utilization of hydoxychloroquine is acknowledged as commonly accepted for patients with malaria and autoimmune diseases, however its utilization where not demonstrated and without clinical management can cause genuine results and ought to be maintained a strategic distance. This research carried out a sentiment analysis regarding the effectiveness of hydroxychloroquine in the treatment of COVID-19. Sentimental Analysis is the way toward recognizing concept from text written based on Natural Language Processing the element it is alluding to. Twitter is an informal community that allows clients to post their suppositions about current issues, share their get-togethers, and associate with others. Twitter has now gotten probably the biggest wellspring of information, with more than 200 million dynamic clients month to month. The technique concentrates and investigations sentimental data from microblogs to forecast the patient's assessment of hydroxychloroquine. In this work, a pre-handling strategy for assessment mining is executed and will be used for examining patients' remarks on Twitte' social media about hydroxychloroquine. The different content pre-handling strategies have been used on the dataset to accomplish a sufficient standard text.

KEY WORDS: COVID-19, HYDROXYCHLOROQUINE, TWITTER DATA, SENTIMENT ANALYSIS, PREPROCESSING.

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic. As of 9 September 2020, more than 27.5 million cases have been reported across 188 countries and territories with more than 897,000 deaths; more than 18.5 million people have recovered (Lai C.C et al., 2020).

This chart shows the number of daily confirmed COVID-19 cases from December 2019 to September 2020. As on 10 September 2020, the total confirmed positive cases have been reported as of 27.89 million in world and 4.47 million in India. There are no specific medicine or proven vaccines for treatment of COVID-19. The executives includes the indications treatment, strong consideration, detachment, and exploratory measures. The World Health Organization (WHO) and the FDA in a Drug Safety Communication declared that hydroxychloroquine has

been given Emergency Use Authorization for the patients treatment who have tested positive with coronavirus (ourworldindata, 2020). Plaquenil and other brands sold Hydroxychloroquine for medication used to hinder and treat wilderness fever in intestinal infection region that stays sensitive to chloroquine.

Figure 1: Total Confirmed COVID-19 cases Across the World and India
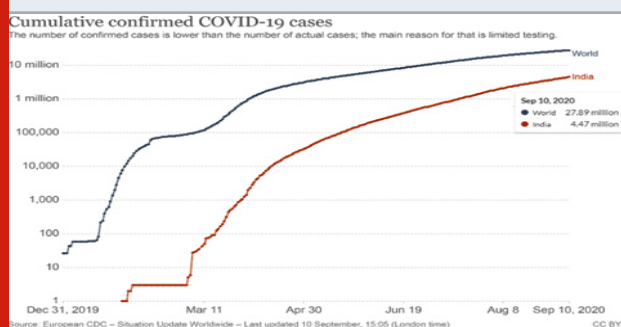
Various treatment uses porphyria cutanea tarda, lupus and joint agony. Hydroxychloroquine sulfate is taken by mouth for the treatment. Hydroxychloroquine is applied to treat and hinder Covid disease 2019 (COVID-19). Basic results may incorporate regurgitating, migraine, changes in vision, and muscle shortcoming. Extreme results may incorporate unfavorably susceptible responses, vision issues, and heart issues. Albeit all danger can't be barred, this consider as a treatment for rheumatic illness during pregnancy (Colson et al. 2020). In United States in 1995, Hydroxychloroquine was supported for clinical use. In WHO Essential Medicines list, it is present. In United States in 2017, it was the 128th most ordinarily suggested medication with more than 5,000,000 arrangements. The theoretical utilization of hydroxychloroquine for COVID-19 compromises its accessibility for individuals with set up signs (Colson et al. 2020).

Motivation and our contributions: In our proposed framework, opinion investigation and AI methods can be used to deal with huge measure of twitter information and cleverly anticipate the impacts of the hydroxychloroquine medicates proficiently. Conventional strategies for announcing drug impacts incorporate clinical preliminaries and unconstrained detailing which has been a successful answer for post-showcasing observation of endorsed meds in typical clinical use, distinguishing numerous significant signs. As of late, mining twitter information has been investigated in distinguishing drugs impacts.

A tweet contains an enormous heap of appraisals about the diverse space which are passed on in various propensities by various clients. It is from time to time basic to standardize the substance for any NLP (Natural Language Processing) attempts. Tweets are frequently tended to in accommodating/unstructured way. Thusly

orderly pre-treatment of tweets is required to refresh the accuracy of appraisal analyzer (Zhao et al., 2017). This paper execute the tweet extraction and the preprocessing stage. Comparable python structure is utilized to finish the preprocessing stage to address the resulting stages feasibly.

Article structure: The rest of the paper is formulated as follows: Literature Survey is present in Section 2, Hydroxychloroquine related adverse events discussions is provided in section 3, Social Media analysis is stated in section 4 and twitter social media outlines is provided in section 5, Sentiment Analysis (SA)/Opinion Mining (OM) details is provide in section 6, methodology explanation is provided in section 7 and conclusion is in section 8.

Literature Survey: Prabhsimran Singh et al. (2017) focused on 'Demonetization' government system based on the twitter data, further applied in politics field. State wise assessment is carried out in this method. The API is applied in the cloud for examination of twitter data and classify into several categories such as "Positive, Negative, Very Negative, Very Positive, and Neutral". Amolik et al. (2016) proposed end assessment for taking films (Hollywood and Bollywood), for medium wellspring and Feature-Vector and classifiers such as SVM and Naïve Bayes for accurately assemble tweets. Naïve Bayesian has better performance compared to SVM for lower survey and precision. SVM has higher accuracy compared to Naïve Bayesian. Precision is increases with increases of the data. Liu et al. (2017) proposed the technique which is set up on the conclusion investigation draws near and the intuitionist fluffy set hypothesis to rank the items through online audits. In that exploration paper, the creator basically centers around the online audits in Chinese. In the technique, online audits of the elective items concerning the highlights are crept utilizing the crawler programming.

Harish Rao et al. (2017) was centered around the examination the viability of the Vader device for solo online estimation extremity investigation. In that work, creator have been effective to utilize the Vader online item audit and solo opinion order. The creator is additionally said that how significant are area explicit preparing datasets to get the necessary precision of grouping, they have attempted to examine this perspective moreover. This examination shows how information representation apparatuses can give more data on an element or viewpoint based assumptions to get significant hints to client estimation identified with explicit highlights.

Aggrawal, Niyati, et al. (2017) expressed as summed up brand prominence examination utilizing system to gauge the WebPages and its position and online media brand page notoriety with identifying with three positioning

components: 1. Alloted rank to WebPages based on site page availability from accessible web indexes and rank connected WebPages of internet searcher to the yield WebPages. Feddah Alhumaidi AlOtaibi et al.,(2019) presented a methodology to perform sentiment analysis by using unsupervised machine learning algorithm. The data here considered is about KFC and McDonald and find which is famous among them. The data that is collected was fed with various models and the output that is obtained is tested using different metrics like cross validation and f-score.

Rasika et al. (2018) suggested different types and techniques to perform sentiment analysis. They defined that there is an increase in accuracy when different opinion mining techniques are used. Prakruthi et al. (2018) proposed that real-time sentiment analysis was done by fetching tweets using Twitter API. Then the pre-processed tweets are compared with the Visualization of Real-time Twitter Data based on Sentiment Classification and then the tweets are labelled as positive, negative and neutral and the outputs are shown by a pie chart. Bhavitha et al. (2017) was centered strategies around a few AI procedures to dissect suppositions and assessment for foreseeing the item surveys. Purchaser keens to know the audit of recently dispatched item and presents a detail item overview of various machines learning procedures. At that point these nitty gritty review was contrasted and precision, favorable circumstances and restrictions of every strategies.

Abinash Tripathy et al. (2014) recommended that estimation investigation is the notable part of characteristic language preparing. To decide the goal of the creator, it manages the content order of text. The expectation might be of gratefulness (positive) or analysis (Negative) type. The method additionally introduced an advancement examination of results acquired through characterization calculations: Support Vector Machine (SVM) and Naive Bayes (NB). These arrangement calculations were utilized for characterization reason in a wistful audit having either a gratefulness (positive) or analysis (Negative) survey.

Geetika et al. analysis AI semantic examination to arrange the sentences and reviews based on the twitter data. The method applies the twitter dataset for tweet analysis. The precision was resolved and improvement of 1.7 % in WordNet semantic assessment was followed up and improve from 89.9 % from 88.2 %. Bac Le et al. (2015) separated twitter data so that twitter is a smaller than usual composition for a blog page. Twitter user may post reviews to the friends. As of now a days, the twitter is a huge dataset of reviews guessed inclines. It presents a technique for picking another rundown of capacities subject to Information Gain, Bigram.

Furthermore, Object arranged extraction procedures proposes an appraisal examination model reliant on Support Vector Machine and Naive Bayes in inclination assessment on relational communication side. The principal explanation behind existing was to dismember determination in a more successful manner. Xing Fang, Justin Zhan, (2017) applied Sentiment analysis on Amazon Online Products extracted from amazon.com. Machine learning algorithms is applied to unmitigated idea limit issues. Random Forest, SVM, Naïve Bayesian are used for various libraries.

Hydroxychloroquine Related Adverse Events: Medication use in medication depends on a harmony between anticipated advantages (as of now explored prior to showcasing approval) and potential dangers (i.e., unfavorable impacts). Clinical pharmacology manages the danger/advantage evaluation of meds as helpful instruments. This should be possible at two levels,

- The individual level, which manages suitable medication solution to a given patient in ordinary clinical consideration and the populace level, which exploits epidemiological apparatuses.
- Systems to acquire answers from past experience.

The two levels are entwined and cover correlative capacities. Existing techniques depend on patients' "unconstrained" self-reports that bear witness to issues. Mining twitter messages makes Pharmacovigilance. In this paper, we portray a way to deal with discover the impacts of Hydroxychloroquine medication by breaking down the substance of twitter messages using assessment mining investigation. To dig Twitter messages for results, the cycle can be isolated into two sections:

- Identifying the Hydroxychloroquine related twitter post
- Finding possible side effects mentioned in the users' Twitter timeline.

Analysis of Social Media: Social media alludes to the methods for collaborations among individuals where they make, offer, and trade data and thoughts in virtual networks a lot on the web. There are various sorts of online media and new types of web-based media are consistently being created and utilized. Web gatherings (additionally called conversation sheets, message sheets, Web discussions, and so forth) are one method of conveying on the Internet. An Internet gathering is a web application that is utilized to distribute client created content under the type of a conversation. Conversations considering specific subjects are called strings or themes. Web discussions have a significant social perspective. Numerous discussions are dynamic for an extensive stretch of time and pull in a gathering of clients that fabricates a network. This accumulation of

the most well-known informal organizations worldwide by dynamic clients (January 2020) arranged by Statistic utilizing information from the Global Web Index board gives an away from of the quantity of dynamic clients (in millions) in Fig.3.

Figure 2: Global social media statistics 2020 in millions
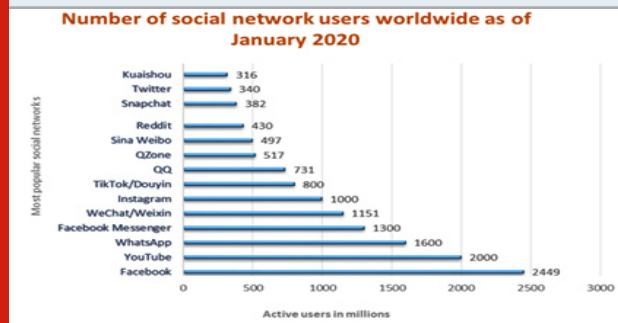
Twitter: Twitter is an American microblogging and long reach relational correspondence organization on which customers post and work together with messages known as "tweets". Selected customers can post, as, and retweet tweets, anyway unregistered customers can simply get them. Customers access Twitter through its site interface, through Short Message Service or its wireless application programming ("application"). Twitter, Inc. is arranged in San Francisco,

Table1. Characteristics of twitter data	
Limited tweet size	It has only 140 characters in hand, so important statements are easily generated that results sparse set of features.
Use of slang	Slang term that alludes to the utilization of twitter to circle news and data with this twitter clients can nonchalantly utilize the articulation and their assessments about explicit theme.
Twitter features	It allows the utilization of hash labels, client reference and URLs. These require distinctive handling in contrast with different words
User variety	The users express their assessments in an alternate manners, some utilizing distinctive language in the middle of, while others utilizing rehashed words or images to communicate their feeling.

Figure 3: Number of users in Twitter social media (2010 – 2019)
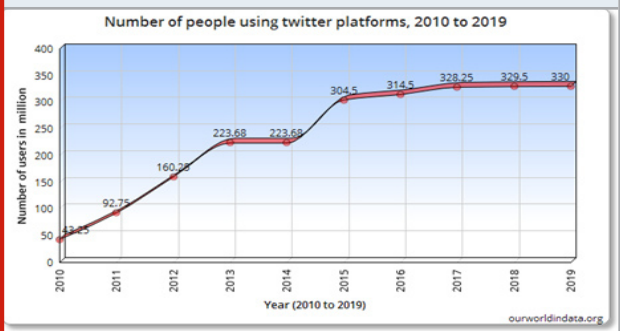
California, and has more than 25 work environments around the world. Tweets were at first restricted to 140 characters, anyway was increased to 280 characters. Sound and video tweets stay limited to 140 seconds for most records (Vishal A, et al., 2016). The point while performing twitter assumption investigation is characterizes the tweets in various feeling classes precisely. In this field of exploration, different procedures have developed, which think of techniques to prepare a model and afterward test it to check its viability. Performing conclusion investigation is trying on twitter tweets. Here we characterize a few purposes behind this: The percentage of adults who use twitter media increased from 2010 to 2019 is illustrated in the below figure.

Twitter can be recognized as one of the biggest person to person communication destinations. An enormous number of clients have acknowledged Twitter as a general stage for getting out the word, imparting insights and associating with others universally. Therefore, a particularly high-volume, high-speed flood of Twitter information produced at each second have the capability of being used for huge insightful and understanding purposes.

Sentiment Analysis: Sentiment analysis or opinion mining is the process to identifying and recognize or categorized the users 'opinion for any services like movies, product issues, events or any attribute is positive, negative or neutral. The sources for this analysis is social communication channels i.e Web site which include reviews, forum discussions, blogs, micro-blogs, Twitter etc. This research field is very popular nowadays because of its opinioned data where user can find reviews for any products/services which are useful for their daily lives. The large amount of opinionated data is stored in digital forms. For particular topic or any opinion the sentiment analysis which relates the mining of the data works and give the output. For sentiment analysis, research works on emotion-based summarization, feeling or mind extraction (Zhao Jianqiang, et al., 2017).

Sentiment analysis also known as Opinion mining which uses NLP – Natural Language Processing to following the emotions, feelings of the public opinion about a particular topic for any product/services. As Sentiment analysis is very famous, it can be also useful in many ways in surveys and advertisement campaign by getting the success rate of any product or services with people's opinion or suggestion. It also give the information about people liking and hating and company gets much clear idea regarding its product features. Sentiment Analysis has increased a lot of acceptance among various zone like politics, business and marketing/selling and advertisement (to estimate sales of specific products). So identifying type of sentence is the most important part of opinion mining. We have to classify the sentence either subjective or objective. Recent or existing research is using both supervised and unsupervised learning technique to provide different techniques for several purpose of sentiment analysis. In initial research all or combination of below supervised techniques are used.

- Support vector machine
- Maximum Entropy
- Naive Bayes

Unsupervised techniques used by initial research are
- Exploit sentiment lexicons
- Grammatical analysis
- syntactic patterns

In general, Sentiment Analysis includes advanced processes. The analysis has a totally different series of tasks, sentiment classification (supervised or unsupervised), subjective or objective analysis, and opinion extraction. For any text document or a sentence, to do the labeling as subjective or objective can be evaluating by subject level analysis. The sentiment classification includes the duty of probing the sentiment polarity of the filtered sentences. All the sentences are dividing in to neutral, negative or positive type which is depends on the opinions we get from the reviews.

1. Subjective / Objective Classification: One of the main tasks in sentiment analysis (SA) are Subjective and objective classification. The main purpose is to divide the entire documents or sentences into one of the two classes; objective or subjective (Zhao Jianqiang, et al., 2017). In Opinion mining, the sentences in the text are labeled as either subjective or objective. By using the sentiment analysis we can extract subjective sentence. The factual information generated by objective sentences need to be removed. When it contains emotions or feelings subjective sentences are important to the procedure with sentiment analysis. Subjective sentence are made up with following views of users, perspectives, thoughts, comments and opinions about the sentence level. Example

- **Subjective:** Inception is an awesome movie. (The sentence having emotion (awesome), thus it is subjective)
- **Objective:** Leonardo is the actor of Inception. (There is no sentiment in this sentence, it is an actual fact, and thus it is objective).

For subjective sentence various research are going on. For any organization, objective lines will be eliminated and subjective sentence will be consider in research as it contain the opinion and emotion. English language is mostly uses for the subjective classification research but other various studies being done in many languages .Arabic language and urdu language are one of them with the use of different supervised learning classification methods and support vector machine (SVM) is one of them for the subjectivity and sentiment analysis (Zhao Jianqiang, et al., 2017).

Resources of sentiment analysis: To collect data is the main purpose of Sentiment analysis where social communication channels like Twitter, Facebook or any pre-existing resources.

1. **Blogs & Forums:** It is source of opinions and emotions where we get information for research purpose and that all information can be used by researchers via Web forums and blogs. Generally, for only single subject forums are designed; thus, by using the forums we can ensure the sentiment mining in single domain. As well, it's the trend that bloggers updated their blogs and reviews every day after activities in and around their areas, countries and around the world.
2. **Reviews:** There are many available studies which dedicated only on reviews because of their usability with the opinions and sentiment. During any research, Movie and product reviews were mostly studied by researcher where the main purpose is to get the feedback from the sentiment and opinions.
3. **News Articles:** News articles, such as financial articles and political reviews are a popular source of sentiment analysis. The main format of News articles texts is structured and formal.
4. **Social Networks:** Many social networks sites are available from which we can take the opinions and reviews for sentiment analysis like Twitter, Facebook, etc.

- Twitter: Tweets are the messages posted by different users, having restriction of 140 characters. Users can read message (called Tweets) of one another. The micro-blogging service which provides this facility is knows as Twitter. By using this tweets which can work as opinions and reviews for future patterns where we can generate the poll results.

- Facebook: The provision of posting personal profile, photos, videos and other related information are provided by most famous social networking facility called Facebook which is popular right after it got launched in 2004.
- Hence, these much ample amount of information available in form of user's message, computer technology which is dependent on sentiment behind this message is introduced known as sentiment analysis (Rashmi H Patil, et al., 2017).

Sentiment Analysis System: Estimation investigation framework is considered as the cycle of text mining. Estimation extremity is the yield of the framework.

Feeling investigation framework comprises of following advances: The work stream for notion examination is appeared in Figure2. The system involves in the four major modules: data collection, data processing, classification and output analysis.

Input (Keyword): The subject is selected and collected the tweets based on tags on the twitter data.

Tweets Retrieval: Tweets can be present in unstructured, semi-structured, and structured data. Tweets can be extracted using R or Python programming language.

Pre-Processing: Information pre-preparing is only separating the information to eliminate the deficient boisterous and conflicting information (Zhao jianqiang, et al., 2015). Following tasks are involved in pre-processing task:

- Removal of Retweets
- Removing URLs, Special characters, Punctuations, Numbers etc.
- Removing Stopwords
- Stemming
- Tokenization

METHODOLOGY

To extricate the assessment, first all information is chosen and separated from twitter as tweets. In the wake of gathering the informational index, these tweets were cleaned from emoji images, superfluous accentuation marks and so forth, and afterward data set is made to store this twitter information in a particular changed structure. In this structure, all the changed tweets are in lowercase letter sets and are secluded into different bits of tweets in the specific field. The insights concerning the means received for the change of data are portrayed in next subsections.

Collecting Hydroxychloroquine Related Tweets: Twitter is a social networking platform by which opinion data is generated continuously. Twitter data were collected through the use of Twitter API (Application Programming Interface) 1.0 in R Tool which only allowed searching for tweets posted recently. The Twitter API provides a streaming API to allow users to obtain real time access to tweets. We continuously queried Twitter with drug name (Hydroxychloroquine), collecting a total of 24,748 tweets related to the Hydroxychloroquine.

Figure 4: Extraction twitter data
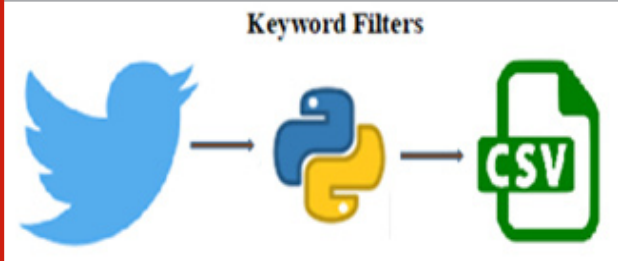
In spite of the fact that this treatment is moderate, it improves the pertinence of the medication related tweets. Impacts are physical or mental signs and conditions appeared on patients who take the prescription. Not all the medication related tweets gathered were identified with drug impacts. This stage includes making a Twitter API and downloading the tweets according to the prerequisites, i.e., downloading tweets of a specific client or tweets having specific catchphrase [24].

Twitter API upholds removing the etymological tweets or the territory based tweets. The information can be recovered in any configuration especially as .txt, .csv, .doc, and so forth, as indicated by the accommodation. We made a Twitter API to gather the tweets. All the tweets related to medicine name and side effects (For example: Metformin and Side effects) were downloaded by providing the keyword "Hydroxychloroquine" in R Tool. The downloaded file was saved in the .csv format. Tweets that describe the author's experience and reactions to the medication were mostly relevant in our study, and are called opinion tweets. These tweets are those that describe the patient's opinions toward the drug. Examples of patient's opinion about tweets are shown in Table 2.

Steps in Pre-Processing: Applying text preprocessing ventures prior to dissecting the tweets is significant for accomplishing the great outcomes [13]. There are a few stages associated with the preprocessing stage, for example,

- **URL Removal:** URLs have nothing to do with assessment investigation. In this way, URLs ought to be taken out from the tweets for viable investigation.

#COVID19 patients treated with #Hydroxychloroquine

and azithromycin recover quickly @CliveFPalmer #coronavirus... https://t.co/gp57Lg4MVw
#COVID19 patients treated with #Hydroxychloroquine and azithromycin recover quickly @CliveFPalmer #coronavirus...

Converting to Lower Case: Upper and lower characters of the tweets text are applied [6]. Twitter data is converted into lower case to process the data.

#COVID19 patients treated with #Hydroxychloroquine and azithromycin recover quickly @CliveFPalmer
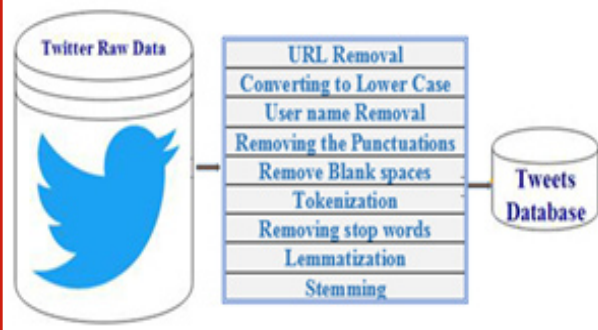
#coronavirus...
#covid19 patients treated with #hydroxychloroquine and azithromycin recover quickly @clivefpalmer #coronavirus...

- **User name Removal:** In Twitter messages, practically sentence may contains a names. The quality doesn't contain any assessment. Along these lines, it is significant advance to eliminate that in pre-preparing step.

Table 2. Sample Twitter data	
Sl.no	Text
1	@weijia @FryGerard @US_FDA It's strange for the FDA to allow #Hydroxychloroquine for emergency use, when HCQ's job... https://t.co/VNw3G0KGvL
2	RT @SIDPharm: New brief report published in @JAMACardio evaluated risk of QTc prolongation associated w/ #hydroxychloroquine +/- concomitan...
3	Dr. Robin Armstrong says he administered #hydroxychloroquine , #azithromycin and #zinc to all patients and monitore... https://t.co/QQfPymjkYR
4	RT @mitchellvii: I'm hearing that blood clots are a major cause of death with COVID-19. This makes sense as COVID-19 is known to cause a cy...
5	RT @mortenoxe: Studying the scientific background for the clinical trials with #HydroxyChloroquine, is like peeling an onion:layer upon la...
6	RT @mitchellvii: #Hydroxychloroquine is intended to be given to people newly infected with COVID-19. The VA study gave it to people alread...
7	RT @MusaShiK7: @Covid19Crusher In Israel EVERY hospitalized patient today is observed for 1 day and if not improved - given #Hydroxychloroq...
8	RT @JimmyFalk5: @gatewaypundit @OneDayOrTheNext Nurse speaking about a New York City Hospital.<U+27A1><U+FE0F> No #Hydroxychloroquine given <U+27A1><U+FE0F> No Respo...
9	RT @drscott_atlanta: #WuhanVirus cases are grossly inflated#FDA prevented community doctors from prescribing #hydroxychloroquine People...
10	#WuhanVirus #CoronaVirus #CCPVirus Will States and Provinces let doctors treat patients with #Hydroxychloroquine an... https://t.co/GqefCQFpTQ
11	RT @vitdoctor: #Hydroxychloroquine https://t.co/p3HaIqsV27. R\T to help get these enrolled. Prophylaxis and early treament trials. First...
12	RT @JimmyFalk5: @gatewaypundit @OneDayOrTheNext Nurse speaking about a New York City Hospital.<U+27A1><U+FE0F> No #Hydroxychloroquine given<U+27A1><U+FE0F> No Respo...
13	RT @drscott_atlanta: #WuhanVirus cases are grossly inflated #FDA prevented community doctors from prescribing #hydroxychloroquine People...
14	#Hydroxychloroquine https://t.co/p3HaIqsV27. R\T to help get these enrolled. Prophylaxis and early treament trial... https://t.co/rh6wiaRrXy
15	RT @alfranken: Now that studies are showing that treating Covid with #Hydroxychloroquine is dangerous and ineffective, Trump is very excite...
16	Any possibility of using a combination of antiviral #Remdesivir and immunosuppressant #Hydroxychloroquine or... https://t.co/j73yeQkkHp
17	Why is @NBCNews @ABC @CBSNews and @CNN Pushhing the "Hydrocholoquine doesn't work" SO HARD? Is it #TDS ? Because o... https://t.co/qCSklfYTF6
18	RT @BeltGold: Italian scientist says she discovered main mechanism behind COVID-19 - The Jerusalem Post #hydroxychloroquine https://t.co/b...
19	#COVID19 patients treated with #Hydroxychloroquine and azithromycin recover quickly @CliveFPalmer #coronavirus... https://t.co/gp57Lg4MVw

Figure 5: Steps in pre-processing

#covid19 patients treated with #hydroxychloroquine and azithromycin recover quickly @clivefpalmer #coronavirus...

#covid19 patients treated with #hydroxychloroquine and azithromycin recover quickly #coronavirus...

Removing the Punctuations (#, @, etc,): Punctuations doesn't share any commitment toward investigating the assessment of an individual. Consequently, they ought to be eliminated to make examination measure simple.

#covid19 patients treated with #hydroxychloroquine and azithromycin recover quickly #coronavirus...

covid19 patients treated with hydroxychloroquine and azithromycin recover quickly coronavirus

Table 3. Examples of preprocessed data

Original	Pre-processed
RT @SIDPharm: New brief report published in @JAMACardio evaluated risk of QTc prolongation associated w/ #hydroxychloroquine +/- concomitanâ€¦	['new', 'brief', 'report', 'publish', 'evalute', 'risk', 'qtc', 'prolong', 'associ', 'hydroxychloroquin', 'concomitan']
RT @mitchellvii: #Hydroxychloroquine is intended to be given to people newly infected with COVID-19. The VA study gave it to people alreadâ€¦	['hydroxychloroquin', 'intend', 'give', 'people', 'new', 'infect', 'covid', 'study', 'gave', 'people', 'alread']
RT @drscott_atlanta: #Wuhan Virus cases are grossly inflated #FDA prevented community doctors from prescribing #hydroxychloroquine	['atlanta', 'wuhanvirus', 'case', 'gross', 'inflatedfda', 'prevent', 'commun', 'doctor', 'prescrib', 'hydroxychloroquin']

Figure 6: Preprocessed data visualization

Remove Blank spaces: This movement is used to wipe out the unfortunate clear space which helps for the tokenization of the tweets.

covid19 patients treated with hydroxychloroquine and azithromycin recover quickly coronavirus

covid19 patients treated with hydroxychloroquine and azithromycin recover quickly coronavirus

Tokenization: Tokenization means breaking the sentence into words.

covid19 patients treated with hydroxychloroquine and azithromycin recover quickly coronavirus

['covid19','patients', 'treated', 'with', 'hydroxychloroquine', 'and', 'azithromycin', 'recover', 'quickly', 'coronavirus']

Removing stop words: This progression meaning to eliminate inflectional endings just and to restore the base or word reference type of a word, which is known as the lemma.

['covid19', 'patients', 'treated', 'with', 'hydroxychloroquine', 'and', 'azithromycin', 'recover', 'quickly', 'coronavirus']

['covid19', 'patients', 'treated', 'hydroxychloroquine', 'azithromycin', 'recover', 'quickly', 'coronavirus']

Lemmatization: This step eliminate inflectional endings and return to word dictionary, which is known as the lemma.

['covid19', 'patients', 'treated', 'hydroxychloroquine', 'azithromycin', 'recover', 'quickly', 'coronavirus']

['covid19', 'patient', 'treated', 'hydroxychloroquine', 'azithromycin', 'recover', 'quickly', 'coronavirus']

Stemming: It alludes to an essential exploratory cycle which cleaves off the closures of words.

['covid19', 'patient', 'treated', 'hydroxychloroquine', 'azithromycin', 'recover', 'quickly', 'coronavirus']

['covid19', 'patient', 'treat', 'hydroxychloroquine', 'azithromycin', 'recover', 'quick', 'coronavirus']

The above table shows the couple of instances of preprocessed twitter information. When the means

are finished, this exploration moves to the following fundamental technique called highlight extraction. Extraction of important words from the tweet is called as highlight extraction. The figure 6 shows representation for assessment of different medications results from patients. This figure provide us knowledge. Indeed, the information started in tweets about hydroxychloroquine are new hotspots for analysis the impacts of clients. A significant problem in this sort of COVID-19 circumstance, this information is ordinarily used to settle on significant choice for some expert, for the most part the thing the client is thinking and confronting the impacts of this medications, However, these assessment can be essential for dissecting the hydroxychloroquine conduct.

CONCLUSION

To separate the assessment, first all information is Coronavirus (COVID-19) is a viral infection brought about by extreme intense respiratory disorder Covid 2 (SARS-CoV-2). Hydroxychloroquine have been discovered to be proficient on SARS-CoV-2, and answered to be effective in COV-19 patients. We assess the impact of hydroxychloroquine utilizing conclusion examination and AI models. Online Social Networks have been progressively embraced by web clients keen on imparting their insights and considerations about any items.

This research has been conducted over hydroxychloroquine medicine related data which originated from Twitter. However, this framework faces part of difficulties in twitter information, because of the casual idea of the posts and the absence of consideration regarding the syntactic guidelines found on client produced content. Here, we experiment with a series of preprocessing methods that applied on twitter dataset for user name removal, punctuations removal, links removal, stop words removal and finally all tweets are converted to lower case. Finally, the raw dataset is then transformed into more useful structured data to improve the classification accuracy.

REFERENCES

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).

Aggrawal N, Ahluwalia A, Khurana P, Arora A. (2017). Brand analysis framework for online marketing: ranking web pages and analyzing popularity of brands on social media. Social Network Analysis and Mining, 7(1)

Bhavitha B.K., Rodrigues, A.P., Chiplunkar, N.N. (2017) Comparative Study of Machine Learning Techniques in Sentimental Analysis. International Conference on Inventive Communication and Computational Technologies (ICICCT), 216-221.

Colson P., Rolain J.M., Lagier J.C., Brouqui P., Raoult D. (2020) Chloroquine and hydroxychloroquine as available weapons to fight COVID-19. Int J Antimicrob Agents.

El Rahman, S.A., AlOtaibi, F.A.m and AlShehri, W.A. (2019). Sentiment analysis of Twitter data, In 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1-4.

Gautam G. and Yadav D., (2014). Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis, International Conference on Contemporary Computing, 437-442

Harish Rao M and Shashikumar D.R. (2017). Automatic Product Review Sentiment Analysis Using Vader And Feature Visualization. International Journal of Computer Science Engineering and Information Technology Research, 7(4), 53-66.

Jianqiang Z., Xiaolin G. (2017) Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis. IEEE-2017

Jianqiang Z. (2015). Pre-processing Boosting Twitter Sentiment Analysis. In 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity),748-753

Lai C.C., Shih T.P., Ko W.C., Tang H.J., Hsueh P.R. (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents. doi: 10.1016/j.ijantimicag.2020.105924.

Liu, Y., Bi, J.W. and Fan, Z.P. (2017) Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. Information Fusion 36 (2017): 149-161.

Le, B., and Nguyen, H., (2015). Twitter Sentiment Analysis Using Machine Learning Tech-niques Springer International Publishing Switzerland pp. 279-289.

Prakruthi, V., Sindhu, D. and Anupama Kumar. S. Real Time Sentiment Analysis of Twitter Posts. In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 29-34

Rashmi H Patil, Siddu P Algur, (2017). Sentiment Analysis by Identifying the Speaker's Polarity in Twitter Data, IEEE Access, 22(5), 2870-9

Singh, P., Sawhney R.S., and Kahlon, K.S. (2017) Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. ICT Express.

Tripathy, A., Agrawal A. and Rath, SK. (2014) Classification of Sentimental Reviews Using Machine Learning Techniques. Procedia Computer Science. 57, 821-829

Vishal A.K, Sonawane S.S. (2016). Sentiment Analysis

of Twitter Data: A Survey of Techniques. ArXiv preprint arXiv:1601.06971. 2016

Wagh R and Punde P (2018) Survey on sentiment analysis using twitter dataset. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 208-211.

Xing F., and Zhan, J. Sentiment analysis using product review data. Journal of Big Data, 2(1):5.

Zhao, J., and Xiaolin, G. (2017) Comparison research on text pre-processing methods on twitter sentiment analysis. IEEE Access, Vol 5, pp 2870-2879, 2017.