International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-8, June 2020

Detecting Twitter's Impact on COVID-19 Pandemic in India

Kusum, Dr. Supriya Panda

ABSTRACT:India has etched a higher place in the economy as a fast growing country with a large population. India is one of the leading Twitter usage countries, with 13.15 million users as ofApril2020[1].A novel coronavirus(COVID-19), which is a pandemic, has been threatening nearly everywhere. This terrible disease started at the end of 2019 from WUHAN in China and is spreading very quickly virtually all over the world. This disease's whistleblower Dr. Li Wenliang also died from coronavirus on Feb 7, 2020.According to the WHO, on 30 January 2020, the outbreak was declared a public health emergency. In response to COVID19 he called for National Unit and Global Solidarity. All the countries in the world are linked with each other due to globalization, the proportion of labor finances migrating economically. In this paper, Twitter reflects the reality of the world. The main issue like signs and symptoms, prevention measures, and medicines which are related to this disease are discussed. Twitter is used for detecting this disease by analyzing data on social media. Nowadays social media sites are very fast and less costly for communication and exchange of information, ideas, and thoughts. This disease is being monitored by Twitter. If there is any delay it will result in a big damage to not only society but also the country. There are two methods:
1. Monitoring system
2. Awareness and alertness

Keywords: Twitter, COVID-19,SVM,Machine Learning, Dynamic

I. INTRODUCTION

Twitter, created by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams in March 2006, launched in July 2006, has gained rapid popularity around the globe[2].It is an online site used for sharing information and making connections with all over the world. Users of Twitter are increasing rapidly day by day. Millions of messages are posted each day. In many applications, twitter serves as a positive important information resource. Boyd et al.(2010) showed activities of communication[3]. Sakaki et al. proposed a probabilistic spatiotemporal model for real-time event detection[4]. Each twitter user was considered as a sensor. For location estimation, Kalman filtering and particle filtering are used. Twitter is a valuable tool.

Revised Manuscript Received on June 30, 2020.
***** Correspondence Author
Kusum*, Ph.D. Scholar Department of CSE,Manav Rachna International Institute of Research and Studies,Faridabad, India E-mail:kusumprerak@gmail.com
Dr. Supriya Panda, Professor,HOD Department of CSE,Manav Rachna International Institute of Research and Studies,Faridabad,India E-mail:supriya.fet@mriu.edu.in

This study shows the phenomenon of detecting one of the very dangerous virus diseases named (COVID-19). Two approaches are described for the prevention of spreading this disease rapidly.

1. Awareness and Alertness in real-time.
2. Warning based on twitter gives positive effect and helps in early-stage detection.

To know the thoughts of the public for awareness and alertness against this disease surveys were the most accurate method but these methods are very costly and time-consuming mostly not suitable for a disease that spread exponentially

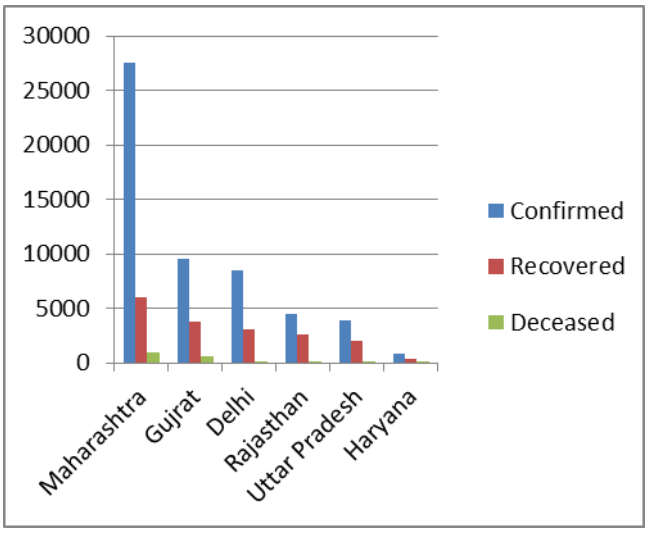
..The best way to solve these problems is to use the Internet, through which information and ideas are easilyshared.Social networking collects unstructured data and shows a user's connections in real-time.

Active Surveillance Service

Many teams are made at the micro level that surveys almost 2000 people's health every day. Flu corners are also made in every state. Dynamic keywords are used to capture tweets from the tweeter. Sentiment analysis stability and hop count-based approach is used on data sets to get potential information towards this disease. This method is also the same as the state of the art method. Basis vocabulary to be used usually chosen based on our own experience or statistics of the corpus.

Longer Monitoring Period

Active surveillance service will prevent huge big loss of lives. Collection of data is done through twitter for 90 days stared from 15thFeb 2020 to 15th May 2020. Monthly cases of COVID-19 deaths, which are cured are shown in the picture:



The issue related to H1NI1, SARS, and other flu diseases also with an examination of most affected states where this disease spreads rapidly are monitored. These techniques prove Twitter as a real-time content, ideas, public alertness, awareness, and attention tracking tool.

According to WHO:

We are the #' United Nations' Health agency. We are committed to achieving better health for everyone everywhere #'Health for all' trending in India.

Active surveillance service investigates those people also which are not positive but are a carrier of this disease who lived with infected persons or have been come in contact with infected people.

While they don't have signs and symptoms of this disorder, they may be a carrier. They are kept in quarantine for 14 days. After 14 days then tests are again taken. Analysis results in an early warning system by awakening and alerting people and informing the method of living can protect from big damage.

A new explorer lab designed to provide complex data sets for easy access and use with variable, selectable across three axes. To help alleviate suffering and save lives who has been working day and night in five key ways:

1.Help build capacity for the country to prepare nd respond.

2. Provide accurate information and fight the infodemic together with numerous partners.

3. Ensuring supplies of essential medical equipment for frontline health workers.

4. Health workers are mobilized and trained

5. Research and development acceleration.

Tweets are very helpful for monitoring public health in real-time. This is also helpful in detecting a state where this disease is spreading rapidly.

II. RELATED WORK:

List of researches has been done in various fields like stock market, election prediction, forecasting and movie revenue and customer reviews ,etc. Limited researches are done in the field of public health by analyzing social media data.

Stewart, et al. suggested early warning system to control outbreak and system analysis[5]. Hansen et al. have done an analysis of big attention to attracting tweets[6]. He presented a piece of evidence that negative sentiments get virality in news columns but not in known news columns. The virality in twitter is based on the probability of retweets. Chew and Eygenbuch gives a method using various meaningful keywords based on twitter during H1N1 pandemic 2009[7]. Elkin et al proposed a machine learning approach(2005)[8]. Chapman et al. proposed an approach based on an algorithm that has been done for finding keywords[9].

Hu et al. made clustering of twitter messages and gave a meaningful label to these clusters by specific terms like flu, (Swine flu), H1N1, etc[10]. Other disease like cholera was investigated by Chunara et al[11]. A framework was developed that provides (to quantify) users affected by this disease within a group or community based on an algorithm. Machine learning technique like SVM was used to predict this disease

Some authors used tools for analysis of twitter data and showed signs and activities of this disease. Chapman et al (2007)applied a machine learning approach. Huang and H.J. Lowe applied novel hybrid approach for automated negation detection in clinical radiology reports[12]

Munjuki Moritadiscussed state of the art method for the NLP technique to extract those tweets that contain useful informatics [13].Samuel et al. proposed a method to provide two classification methods for the classification of coronavirus tweets of varying lengths [14]. The accuracy of classification for short tweets was strong. They used the SVM+Naïve Bayes method.

Lampos and Cristianini used Twitters as a monitoring tool [15]. The main tweets are analyzed per day. Data with Flu score is compared with health protection agency data. Independent data is used in this method. Ginsberg et al. proposed search queries were utilized to detect an epidemic in areas with a large population [16]. For epidemic surveillance, a comprehensive model is generated. Linear regression with cross-validation of for fold was used.

Jain et al. used twitters to make a good surveillance system [17]. Polgreen et al. proposed that search terms surveillance may provide an additional tool for surveillance of disease[18]. Daily unique queries were counted.

Government Health Agencies:

At this critical time when a horrifying disease causes big damage to human life in the world became a need for health agencies and government communications by sharing accurate and relevant information to the public on time by using social media sites or any other media like broadcasting etc.

To overcome the big damage and saving people's lives have become the main issue.

III. PROPOSED METHOD:

This paper addresses pandemic COVID-19 in India during the year 2020.

Keywords in early warning system jargon are used Toobtain awareness of this disease and public ideas about it.

Analysis of public opinions, their information, and government facilities to overcome the disease are investigated. Analysis for the service facilities like labs for testing, availability of medicines, and extra paramedical staff appointed by the government are made.

Data Collection Methodology:

The method adopted is an analysis of tweets and web queries tweets with relevant and meaningful keywords that are identified with relevant and meaningfulkeywords.Significant keywords mean terms that a re used mainly in top popular newspapers.

Main headlines about this topic are collected in the popular newspapers for a definite period.

At this particular period, keywords are prepared ,which are related to public thoughts.

It's based on the weight of the term, which how many times a word is used in a document.

February	March
Outbreak	COVID-19
Novel	Pandemic
Deadly	Distancing
Wuhan	Coronavirus
Spread	Self-isolate

China	Lock-down
Crisis	Self-Isolation
Strain	Sanitizer
Case	Quarantine
Infection	Ventilator
New	Nonessential
Confirm	Self-Quarantine
Epidemic	Virus
Originate	Outbreak
Symptom	Corona
New	Disease
Fear	Postpone
Cause	Disinfect
Spreading	Isolation
Declare	Concern

Analysis of COVID -19 related top keyword data from twitter data collection was performed at different time intervals.

TF-------No. of times words appeared in document
Total no of words in a document

IDF----------Total No of documents
No of documents with words in it

Living System introduced by Government

1. Wash your hands properly after 20 minutes, and where possible, with soap and detergent.
2. Stay at home
3. Do not go outside without any essential work
4. Maintain the distance in the home also
5. Wear masks when you go outside.
6. If you touch any metallic surface then sanitize your hands.
7. Pray for Health care workers.
8. If you see a person sneezing with coughing and fever inform the government.

Collection of data sets

We collected the related keywords from 15thFeb to 15th May 2020 using relevant keywords. Keywords related to signs and symptoms prevention and related keywords.

Signs and symptoms:

Running nose, throat pain, cough, trouble in respiration, a problem in breathing, early symptoms are like SARS but after it becomes severe.

Prevention Steps

Living by the goodness of government
Increase immunity by taking balanced foods.
Stay home in the lockdown period.
Prevention is the only treatment of this disease
Avoid touching hands to mouth, nose, and eyes.
Sanitize your hands.

Medicines:

Tulsi, Ginger, Amla, Giloy, Neem and medicine like HCQ and medicines given in SAARS disease, etc.

Tweets Classification

Tweets are pre-processed by stopword elimination, stemming etc.Then, classification is achieved by differentiating from irrelevant for specific tweets. Every relevant tweet related to COVID-19 is taken as feature and classification is done by SVM classifier by comparing the performance of different classifiers like Naïve Bayes and Decision tree. Keywords used as a tool for detecting targeted COVID19 event tweeters, etc .

We prepare three feature groups A, B, and C,
Features of statistics----------A
Features of Keywords----------B
Control word features----------C

An algorithm calledcontext is created and the condition is negated or encountered by another patient is taken into account.The condition must be known either acute or chronic, or hypothetically mentioned.

These are the features that are used as context. Active Surveillance system and awareness and alertness are imperative issues in today's time, state of art surveillance systems provide limited information. ED reports are a timely source of information and are useful for syndromic surveillance. A natural language application called an active surveillance system has developed. It uses information if the patient has acute URC and fever. Another condition is chronic and the third category is absent. The tweets are collected, cleaned, and prepared. Prepared tweets were downloaded using twitter API.20000 tweets month-wise from Feb to March, March to April, and April to May 2020 with features used as keywords were collected. For analyzing data meaningful keywords related to COVID-19 from different datasets during time intervals were collected. Preprocessing was done to differentiate relevant and irrelevant tweets. Every relevant word related to this disease is considered as a feature. Cleaning process like stop words removal etc. was done. Data sets were further evaluated for the identification of useful variables. Deleting irrelevant values, the cleaned data set was created. Tokenization was done for converting text to relevant words. Part of speech tagging, parsing, stemming and lemmatization was done for the transformation of words to simple forms.//For determining the performance Analysis was compared by different classifiers like SVM, Naïve Bayes, decision tree, etc. with F-measure. There are 4 outcomes with a given classifier: True Positive, False Positive, True Negative, False Negative. The number of positive =True positive +False Negative. The number of negative=True Negative+ False Positive .As a result Number of positive=True 1, Number of Negative=-True-1. F-score is a metric for accuracy. Both recall and precision are calculated. True Positive rate also called sensitivity or recall is denoted by tpr and false-negative rate by fnr. True negative rate=1-false positive rate. false-negative rate=1-true positive rate.

Precision=total positive
Total positive +false positive

Recall=Total positive
Total positive +False negative

F measure=2*precision*recall
Precision +recall

Result was obtained by the equal sets of features

Both fear sentiment, negative sentiments showing relatively level in states were examined. Machine learning techniques, SVM, Logistic Regression method to train test sentiment categories are used. Positive and negative sentiment with sentiment scores is split into test and trained data.

Logistic Regression:: probabilistic method of classification. Machine learning model contains the following functions:
1.Input as a feature representation of a feature vector [x].
2. Function classification or sigmoid function is used.
3. The job of objective function is to minimize the error.
4.Optimizing algorithm: The stochastic gradient algorithm is popularly used for optimizing an objective function.
A binary classifier is made by using logistic regression and sigmoid function classifier output maybe 1 or 0.
The objective is to know P(y=1/x), which tells the probability of negative sentiment.
P(y=0/x) gives the probability of negative sentiments.
W denotes the weight of input features with b denotes bias terms resulted in the weighted sum for class:
z=wx+b
using sigmoid function to map the real values

$$y = \sigma(z) = 1/1 + e \wedge - z$$

$$P(y = 1|x) = \sigma(w.x + b) = \frac{1}{1 + e^{-(w.x+b)}}$$

$$P(y = 0|x) = 1 - P(y = 1|x) = \frac{e^{-(w.x+b)}}{1 + e^{-(w.x+b)}}$$

Data subsets were generated by increasing tweet lengths. Tw o data groups were formed,as
Tweets with less than 77 characters and between 77 and 125 characters. Logistic expressions were used to describe them. The classification text's sensitivity and after that, the text's s pecificity was determined, which gives a positive prediction ratio.
The classification specificity gives a ratio of the number of n egative predictions.
Logistic regression with balanced
tweets is showing better results. This research has thus provi ded a system of useful knowledge and public opinion that can be used to establish the solution and strategies
to combat COVID-19's rapid spread.

IV. CONCLUSION:

Twitter as a kind of communication tool provides specific challenges and opportunities for tracking public health and active surveillance services. From 15 Feb 2020 to 15 May 2020, a health monitoring scheme is proposed to track the crisis in India. This method gives easy knowledge and ideas of living through awareness and alertness. Prevention steps to be taken with official warnings.
This method gives easy knowledge and ideas of living through awareness and alertness. Prevention steps to be taken with government alerts to collect policy and health agency information. Twitter offers rapid details such as signs and symptoms of Covid-19, the technological method of prevention used, public knowledge.
Monitoring health scheme is based on relevant keywords. Analysis of tweets and classification into relevant or irrelevant is done. Twitter is used as a corpus for training classifiers. In revealing situational awareness during crisis scenarios Twitter is useful. In this research a lot more needs to be done across social media, news and public communication platforms to better understand public sentiments and expectations. The solution is very critical in identifying the pathway to recover post-COVID-19.

REFERENCES:

1. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/
2. https://en.wikipedia.org/wiki/Twitter
3. Boyd, D., Golder, S. and Lotan, G., 2010, January. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *2010 43rd Hawaii International Conference on System Sciences* (pp. 1-10).IEEE.
4. Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, April. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (pp. 851-860).
5. Diaz-Aviles, E., Stewart, A., Velasco, E., Denecke, K. and Nejdl, W., 2012, May. Epidemic Intelligence for the Crowd, by the Crowd.In *Sixth International AAAI Conference on Weblogs and Social Media*.
6. Hansen, L.K., Arvidsson, A., Nielsen, F.Å., Colleoni, E. and Etter, M., 2011. Good friends, bad news-affect and virality in twitter. In *Future information technology* (pp. 34-43).Springer, Berlin, Heidelberg.
7. Chew, C. and Eysenbach, G., 2010. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one*, *5*(11).
8. Elkin, P.L., Brown, S.H., Bauer, B.A., Husser, C.S., Carruth, W., Bergstrom, L.R. and Wahner-Roedler, D.L., 2005. A controlled trial of automated classification of negation from clinical notes. *BMC medical informatics and decision making*, *5*(1), p.13.
9. Chapman, W., Dowling, J. and Chu, D., 2007, June.ConText: An algorithm for identifying contextual features from clinical text. In *Biological, translational, and clinical language processing* (pp. 81-88).
10. Hu, X., Tang, L. and Liu, H., 2011, October.Enhancing accessibility of microblogging messages using semantic knowledge.In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 2465-2468).
11. Chunara, R., Andrews, J.R. and Brownstein, J.S., 2012. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American journal of tropical medicine and hygiene*, *86*(1), pp.39-45.
12. Huang, Y. and Lowe, H.J., 2007. A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American medical informatics association*, *14*(3), pp.304-311.
13. Aramaki, E., Maskawa, S. and Morita, M., 2011, July. Twitter catches the flu: detecting influenza epidemics using Twitter. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1568-1576).Association for Computational Linguistics.
14. Samuel, J., Ali, G.G., Rahman, M., Esawi, E. and Samuel, Y., 2020.COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Nawaz and Rahman, Md. Mokhlesur and Esawi, Ek and Samuel, Yana, COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification (April 19, 2020)*.
15. Lampos, V. and Cristianini, N., 2010, June.Tracking the flu pandemic by monitoring the social web.In *2010 2nd international workshop on cognitive information processing* (pp. 411-416).IEEE.
16. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), pp.1012-1014.
17. Jain, V.K. and Kumar, S., 2015.An effective approach to track levels of influenza-A (H1N1) pandemic in India using twitter. *Procedia Computer Science*, *70*, pp.801-807.
18. Polgreen, P.M., Chen, Y., Pennock, D.M., Nelson, F.D. and Weinstein, R.A., 2008.Using internet searches for influenza surveillance. *Clinical infectious diseases*, *47*(11), pp.1443-1448.

AUTHORS PROFILE

Kusum Mehta received her B.E. in Computer Science and Engineering in 2002 from Vaish College of Engineering ,Maharishi Dayanand University Rohtak and received her Master's Degree in Computer Science and Engineering from Banasthali Vidyapeeth in the year 2006. She is pursuing her PhD in Department of Computer Science and Engineering, at Manav Rachna International Institute Of Research and studies, Faridabad, India. Currently, she is doing her research in the area of Big Data under the guidance of Dr. Supriya Panda.

Dr. Supriya Panda received her Doctorate in Computer Science in the year 1990. She has around Thirty one years of Academic experience in the field of Computer Science. Currently She is working as the Professor in the Department of Computer Science and Engineering at Manav Rachna International Institute Of Research and Studies,, Faridabad, India. She has been the best Teaching Fellow at BGSU,Ohio , USA during MS and Ph.D(1985-90).