EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Opinion Mining for the Tweets in Healthcare Sector using Fuzzy Association Rule

Mamta Mittal[1], Iqbaldeep Kaur[2], Subhash Chandra Pandey[3], Amit Verma[2], Lalit Mohan Goyal[4,*]

[1]Department of Computer Science & Engineering, G.B. Pant Govt. Engg. College, Okhla, New Delhi, India
[2]Department of Computer Science & Engineering, Chandigarh Group of Colleges, Mohali, India
[3]Computer Science & Engineering Department, Birla Institute of Technology, Mesra, Ranchi(Patna Campus) Patna, Bihar, India
[4]Department of CE, J.C. Bose University of Science & Technology, YMCA, Faridabad, India

Abstract

Communication among several internet users has become more convenient through social networking sites to where each user sharing his own opinions on different matters, such as Healthcare, Education, marketing etc. The Objective of this paper is to present a method to make it easier for even a layman to predict and analyze one's health issues on his own by making use of tweets on the social website twitter.com. As far as methodology or techniques is concerned, an algorithm has been framed for the same to perform the analysis on health care tweets with association rules to classify the ailments and their symptoms using a corpus through fuzzy set and two step approach for Document Term Matrix & Term Document Matrix. The results demonstrate the comparison of different terms over the WordCloud which concludes that in this novel approach of two step authentication the average accuracy of association between the hiv ailments is 98% through correlation table and association between the HIV ailments with 98% correlation.

Keywords: Health, Social Media Analysis, Twitter Mining, Sentimental Analysis, Text Mining, Association Rule.

Received on 15 July 2018, accepted on 13 August 2018, published on 30 October 2018

doi: 10.4108/eai.13-7-2018.159861

*Corresponding author. lalitgoyal78@gmail.com

1. Introduction

As Internet Based social media has transformed the whole world into a hamlet which has totally changed the meaning of traditional methods of communication. It permits the commencement and communion of different aspects of User Generated Content and their opinions. Thus, in the contemporary era, the social networking sites and media are thriving increasingly in every walk of life. The recent surveys estimated by the Centre of Pew Research [1] suggest the frequency of internet users in India is 65% which are used Facebook and Twitter social networking sites for sharing their interests and effects of life. The Twitter most widely used micro blogger has been used to keep track of the plethora of public health issues [2, 3] and their correlated tweets, such as epidemics, medications, manifestations and alternative therapies [4, 5]. According to Pew Internet & American Life Projects, the most searched area is health-related information during web-linked actions over the social networking sites [6]. Developing countries like India have to face various disasters and epidemics every year and people tweet about diseases very often. Thus, the relationship between various symptoms can be circumspect to measure the scale of epidemics.

1.1 Related Work

Healthcare tweets can be taken as a primary data source for data analysis to detect and monitor [7] the plenty of information about the health of independent users' experiences and health groups decisions [8, 9]. They can help in the timely diagnosis of an epidemic to make people aware and attentive about ongoing effects with healthcare [10]. The current technique works to extract the highly

occurring disease in people and to help in taking the preventive measures by people [33-36].

The current approach reduces the critical element of time to diagnose and treat mild, acute and common chronic conditions by suggesting healthy lifestyle and effective diet regimen. The novelty of current approach lies in additional implementation of the association rule for finding the correlation between the terms specifying with some limits. The Association rules will be applied on the extracted tweets and can be used to draw inferences for the relationship between symptoms of popular diseases like dengue.

The Twitter analysis method is improvised with the Opinion Mining which is current arena of research in text mining [11, 12]. Opinion Mining classified the heath tweets [13, 14] on the basis of scores into positive, negative, extremely positive, extremely negative and neutral sentiment [15, 16] where positivity indicates confidence. Mining opinions from twitter has been a tactical thing where people can express their feeling with honestly about different facets in a liberal way [17] and short text.

1.2 Objective

Opinion Mining design involuntary systems to influence the human beliefs, feelings, thoughts and ideas from a text written in natural languages [18]. In this paper, objectives have been layout as :

- To find association and opinion in the cluster of words
- To analyse and perform techniques for extracting knowledge based results for health mining data

an approach has been used 45,973 tweets for find the association and opinion in an accumulated collection of health care tweets which are plain natural language sentences.

2. Two Step Sentiment Classification Approach

Several authors have worked on opinion mining by using different techniques [19]. The twitter analysis has also been integrated with the sentimental classifications for measuring the public health concerns by using the machine learning approach. [2] address and monitor the problem of public health concerns by using the MEASURE OF CONCERNS (MOC), which is derived from Twitter.

They also develop a dual pace approach to classify the sentiments for deriving the MOC and integrate with the clue-based search and ML methods. In which clue-based Approach uses the Multi-Perspective Question Answering dictionary News stop word list , Profanity words and Emoticon lists .

Two-step Sentiment Classification Approach is as follows (Classification with automation through learning and further classification with authentication through clue based strategy):

• STEP 1: PERSONAL TWEET CLASSIFICATION
• STEP 1a: Raw data as INPUT then Preprocess it
• STEP 1b: After Preprocessing the CLUE-BASED METHOD automatic labeling datasets for training classifier compared with MPQA Dictionary, News stop words, and profanity list.
• STEP 1c: In MACHINE LEARNING BASED METHOD the tweets are labeled as news tweet and personal tweets then build a classifier
• STEP 1d: After classifier trained and classify the tweets of personal and news as OUTPUT.
• STEP 2: SENTIMENT CLASSIFICATION
• STEP 2a: Personal Tweets as INPUT.
•STEP 2b: In CLUE-BASED METHOD the automatic labeling tweets for training classifier compared with Emoticons and profanity list for generate training dataset.
• STEP 2c: In MACHINE LEARNING BASED METHOD the tweets are labeled as Personal neg. tweets and personal Non-neg. tweets then build a classifier
• STEP 2d: After classifier trained and classify the personal neg. tweets and Personal Non-neg. tweets as OUTPUT.
• Afterward the step 1 and step 2, the raw tweets T converted into sequence of label tweet

The appraisal of the dual pace method is founded on two groups of test datasets which are Clue based annotation and Human annotation. The results of the both annotations are evaluated by comparing the three classifiers: NB, Multinomial NB and SVM [22](Generally for SVM- for a given set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible).

Id	P_{CL}	S1				S2			
		T/p	F/n	F/p	T/n	T/p	F/n	F/p	T/n
E_{MC}	S2-SV	047916	09	018	0669	017046	0245	0398	02652
M_{HL}	S2-SV	020602	06	03	0113	04290	069	088	01353
C_{SC}	S2-SV	03571	019	01	0575	0318	005	003	076

Table 1 Confusion Matrices in Clue based annotation on particular domain

Where T= True, F= False, S2-SV= Step 2 Support Vector Machine, E_{MC}= Epidemic, M_{HL}= Mental Health, C_{SC}=Clinical Science, P_{CL}=Paramount Classifier, p=pos., n=neg.

The confusion matrices of the paramount classifier on particular area in clue-based annotation for personal tweets, news tweets, personal neg. tweets, personal pos. and neutral tweets are shown in Table 1. The paramount classifier of the clue-based annotation is SVM classifier on the individual datasets[36]. The clue-based method result produced an assured bias; to overwhelm this difficulty the researcher spawned a succeeding test dataset by human annotation. The confusion matrices of the paramount classifier on collection of Human Annotated data are shown in Table 2.

Table 2. Confusion Matrices on human annotation

Id	P_C	T/p	T/p	F/n	T/n
E_{MC}	**S2-NB**	052	015	011	0122
M_{HL}	**S2-NB**	081	023	021	075
C_{SC}	**Clue-based**	021	029	07	0143

It achieves the best accuracy results using the NB paramount classifier for specific Health area. There are some gaps with the existent proposed method which are generates general hurdles for research.

- To classify the health-related tweets using two-step sentiment classification approach is long-delayed and tedious procedure.
- Continuous loss: Copious of the tweets are eliminated by the current approach which is in the heavy strength and their tweets are collected using the limited keywords.
- Misclassified: Some of the tweets are failed to classify and monitor the health-related personal tweets and news tweets. There are 27 out of 140 errors are found for the time of classifying the personal tweets rather than news tweets and 48 out of 140 errors are found where actual the tweet is related from news but is categorized as a personal tweet.

In this current research, the author implements only to classify and find sentiments [23-24]of the personal and non-personal tweets for obtaining the ailments frequency rather than finding the correlation between the epidemic issues.

Since Zadeh's invention of the concept of fuzzy sets it has been extensively investigate in mathematics science and engineering.

Many authors [11,25,7,3,4]studied the existence of fixed point in fuzzy settings . In the present note we first establish the existence of a maximal fixed point of fuzzy monotone multifunctions . In the second time we prove the existence of fixed points of fuzzy monotone multifunctions by using iteration method.

Let $\boldsymbol{M}$ be a fuzzy memory space of opinion consisting of different disease, with generic disease of $\boldsymbol{M}$ denoted by $\boldsymbol{d}$. A fuzzy sub-memory space $\boldsymbol{S}$ of $\boldsymbol{M}$ is characterized by its membership function

$$\mu_S: \boldsymbol{M} \rightarrow [\mathbf{0}, \mathbf{1}]$$

Equation 1

and $\mu_S(d)$ is interpreted as the degree of membership of disease d in fuzzy sub-memory space S of M for every $d \in M$. Let S and S^*be two fuzzy sub-memory space of M : We say that S is included in S^* and we can write $S \subseteq S^*$ if $\mu_S(d) \leq \mu_{S^*}(d)$, for all $d \in M$. In particular, if $d \in M$ and S is a fuzzy sub-memory space of opinion in M, then $\{d\} \subseteq S$ if $\mu_S(d) = 1$. It is intuitive that opinion mining renders the fuzzy order relation. It is due to the fact that a fuzzy order relation on M is a fuzzy subset R of $M \times M$ satisfying the following three properties:

(i) $\forall d \in M, r(d,d) \in [0,1]$ (Reflexivity).

Equation 2

(ii) $\forall d, d_1 \in M, r(d,d_1) + r(d_{1,d}) > 1$. It implies $d = d_1$ (Antisymmetry).

Equation 3

(iii) $\forall d, d_1, d_2 \in M^3, [r(d,d_1) \geq + r(d_{1,}, d)$ and $r(d_1, d_2) \geq r(d_2, d_1)]$. It implies $r(d,d_2) \geq r(d_{2,}d)$ (f −Transitivity).

Equation 4

Further, it is intuitive to consider the sub-memory space S as a fuzzy chain because the fuzzy ordered relation of disease is said to be total if $\forall\, d \neq d_1$ implicates either $r(d,d_1) > r(d_1,d)$ or $r(d_1,d) > r(d,d_1)$.

Equation 5

Let S be a fuzzy sub-memory space of M. It is natural to say that a disease $d \in M$ will be an upper bound of S if $r(d_1,d) \geq r(d,d_1)\ \forall d_1 \in S$. Further, if disease d is an upper bound of S and $d \in S$ then the disease d will be the greatest element of S. Furthermore, the disease $d \in S$ is called a maximal element of S if there is no $d_1 \neq d$ in S for which $r(d,d_1) \geq r(d_1,d)$. similarly, we can define lower bound, minimal and least disease elements of S.

Moreover, it is obvious that the supremum of fuzzy sub-memory space S i.e., $sup\,(S)$ will be the least element of upper bound of S (if it exists). Similarly, infimum i.e., $\inf(S)$ will be the greatest element of lower bound of S (if it exists).

Further, in this sequel we can say that if we consider M as fuzzy order set then it can generate a map of fuzzy multifunction $T: M \rightarrow [0,1]^M/\{\phi\}$ such that for every disease $d \in M, T(d)$ is a nonempty fuzzy sub-memory space of M. Furthermore, the fuzzy multifunction $T: M \rightarrow [0,1]^M/\{\phi\}$ is said to be fuzzy monotone if and only if

$\forall\, d, d_1 \in M, r(d,d_1) \geq r(d_1,d)$

Equation 6

and it implicates that for all

$\{a\} \subseteq T(d)\ \exists \{b\} \in T(d_1)$ such that $r(a,b) \geq r(b,a)$

Equation 7

3. Methodology

In this section, the collection of micro-blog datasets has been described and a novel method has been presented that processes the extracted datasets to obtain the multiple decisions on the basis of different possibilities. The current approach has been implemented using R statistical tool and Twitter[20-21] APIs version 1.1 integrated with relevant packages like tm, twitter and word cloud which are available on CRAN to acquire the healthcare tweets and allow for speedy visual analysis of keywords that originate from the health-related tweets.

This tool fetches the real time data by tracking the latest healthcare trends containing the specific hashtags. The current approach is distributed in various steps. Every single step is independent of the others but significant at the same time.

3.1 Data Collection

The dataset used has been obtained through the Twitter Application Programming Interface (API) Streaming approach which can be further accessed with Register Twitter Open Authentication (OAuth) to retrieve the relevant tweets based on health keywords and hashtags.
It is easy to crawl and collect the regular feed of data from Twitter.com. To ensure that we load the tweets into R and then redirect all the tweets to csv files that refers with healthcare ailments and their symptoms.
There have been extracted 45,973 tweets over the period of last three months. The collected data is used to convert into the data frames and build a corpus to mine the most frequent words. The Random Sample health tweets are mentioned in Table 3.

Table 3. Random Sample Tweets

Postdate	Content
11/2/2016	#Malaria is a disease transmitted by mosquitoes that bite at night.
11/2/2016	Did you remember to take your #hiv pill? We hope you did! And far away #HIV
11/2/2016	RT @Malaria_Mission: Our @ThunderclapIt has been supported ro 109%! And we still have 21 hours left BIG thank you to all supporters!#nausea
11/2/2016	RT @Vashishtrv: Single-cell genomics of Malaria @GenomeResearch #SingleCell #SystemsBiology #Mathematics #Disease
11/2/2016	#Malaria Put on your mozzyproofs: Jaeger founder gets the fashion bug again - making clothes to help fight malaria... #nausea #vomiting
11/2/2016	You can't be vaccinated against #malaria but you can protect yourself by Avoiding Bits and Taking Antimalarial Tablets
11/2/2016	#Malaria kills 55 in Tripura: 46 children, have died in malaria since the first week #kidneyfailure #coma #yellowskin #sweating
11/2/2016	RT @sassafras4ucom: 9 (6-1): #Protein that may lead to #malaria #vaccine discovered
11/2/2016	RT @Floramujaasi: #HealthFocus @ntvuganda in the last of the series on #malaria (not last story though) we look at the reduction in deaths

All the recent tweets are retrieved in English language during the collection phase and all the content of the tweets is accumulated as a corpus from exported files to execute the preprocessing step.

3.2 Pre-Processing

Preprocessing is one of the important phase of data mining task. The preliminary cleaning is covered in preprocessing step by transformations [26, 27, 31, 32]. The Tweets are required to transform formerly mining the meaningful info by eliminating the noisy text. Twitter data alsocontains the slangs and replicated characters which requisite to be uninvolved. Initially the tweets have been removed the 'RT' keyword which specifies the retweet forwarded by another individual user since it just represents redundancy. For the remaining tweets the punctuations, stop words, numbers, white space and special characters are removed.

After that the URL of the tweets are substituted by string "url". The 'tolower' content transformer function used to convert the upper case into lower case of corpus. Then replace the number of patterns like "treatment" in place of treat, treatmented and treatmentments, HIV instead of hivaid and hivs, test inplace of tested and testing, cause in place of causes and caused, prevent instead of prevention and prevented and likewise.

3.3 Build DTM and TDM

Document Term Matrix is created to obtain the lists of all the occurred health-related words of corpus into the matrix where terms are considered in rows and occurrence are considered as columns.
On another side the Term Document Matrix transposes the matrix as opposed to DTM where column becomes the rows and rows become the columns. In Table 4 the summary information on the matrix is given.

Table 4. Evaluation of terms and documents

<<DocumentTermMatrix (documents: 4, terms: 42787) >>	
Non- /sparse entries	46188/124960
Sparsity	73%
Maximal term length	54
Weighting	Term frequency (tf)

As we can see from the above results, document-term matrix is composed as a 4 × 42787-dimension matrix in which 73% of the rows are zero.

3.4 Mining Corpus targeting Keywords

TDM converts a corpus into a measuring object that can be examined using quantitative techniques of matrix.

- The Mining corpus is to identify the keywords by matrix algebra.
- The term occurrences are sorted into descending order and hence it gives the most and least frequent targeting terms.
- After that the mathematical matrix is converted into the word frame which is shown in given below Table 5.

Table 5. Extremely Occurred Terms

Term	Occurrences
Hiv	34848
Malaria	11090
Aids	8462
Test	6418
Diarrhea	6150
Vomiting	5958
Nausea	5867
Headache	5724
Sweating	5717
Anemia	5701

Where the HIV ailment is more frequently occurred in health-related corpus then followed by malaria then aids and it goes sequentially in descending order.

4. Experimental Results

In this fragment, this paper demonstrates the research and appraises the results using the above methodology. To evaluate result from the huge number of real time tweets acquired from twitter is very challenging. Thus, the results of the health-related tweets in the current work have been obtained with the association rules and opinion mining. The proposed approach has been implemented with the use of R statistical tool and Twitter API's. The Implemented method is integrating with the R packages which are available on the Comprehensive R-Archive Network (CRAN).

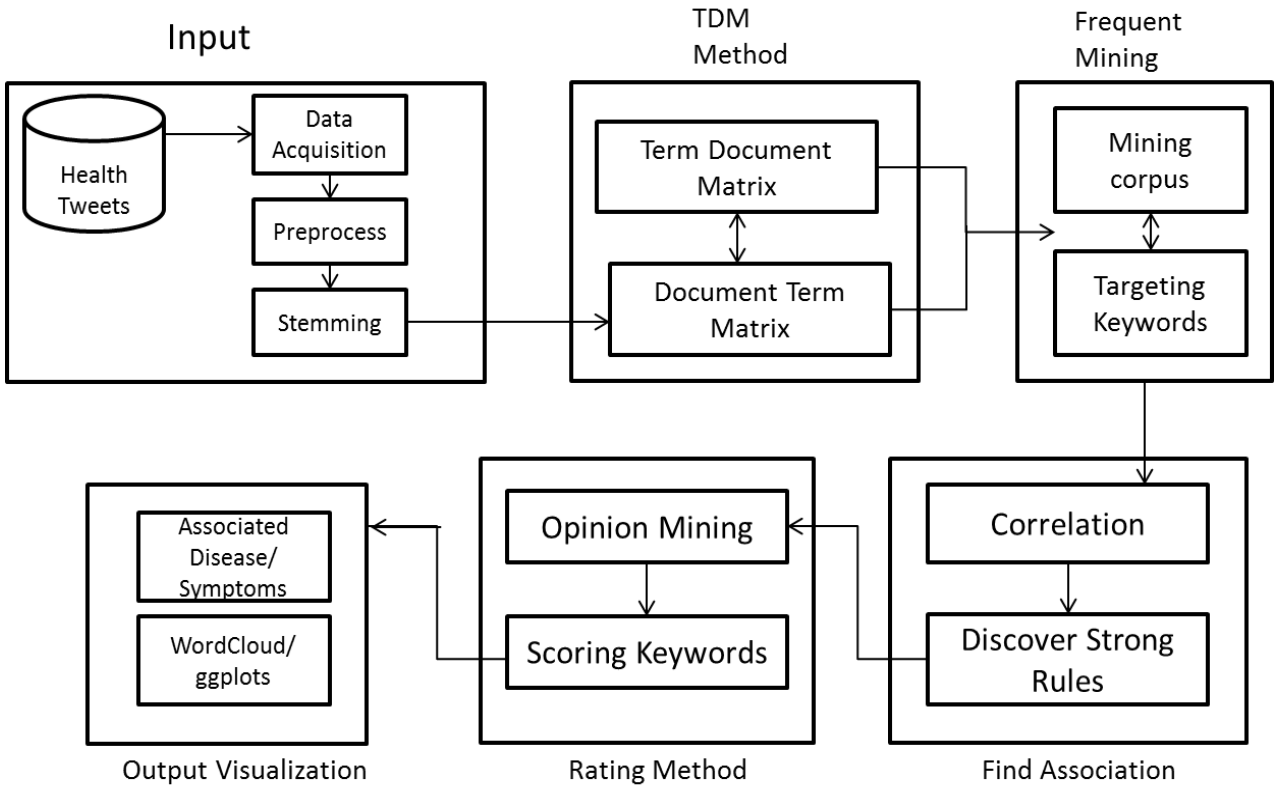


Figure 1. Flow of the proposed Work using Fuzzy Association Rules

It is used to predict the extremely occurred health issues from the social media platform on the basis of preliminary findings those are mentioned in Figure 1 below. Further the figure below provides the process and working with various phases as Novel Approach.

4.1 Frequent Terms and Association

The occurrences of the health terms are evaluated from the corpus by using the ggplot2 package. It also determines the concerning relationships using the association rules on the basis of correlation limits.

The often-occurred terms of corpus are arranged as alphabetically and graphically which is demonstrated using the word frame in Fig 1. The illustration of most frequent terms is:

- "HIV"
- "malaria"
- "aids"
- "test"
- "sweating"
- "vomiting"

The limit of correlation is a number between 0 and 1 which serves as a lower bound to upper bound for the search and outcome terms. Table 6 and Table 9 shows the results of operating the association function on sweating, diarrhea and HIV which are frequently occurring terms at a correlation of 98%.

The results of common terms as symptoms which are most commonly associated with ailments on the basis of their correlation limit values range from 0.0 to 1.00 are summarized in Table 7.

4.2. Opinion Mining

Applying the opinion mining on the healthcare tweets can be used for finding the opinions of the individual users based with several categories such that pos., neg. and neutral sentiments.

In Table 8, the keywords of corpus are examined using the scale between -5 which is enormously negative to +5 which is enormously positive on the basis of negative and positive dictionary [28]. If the tweet score is equal to zero then the keyword is neither positive nor negative called neutral sentiment.

The scoring manner commonly improved the scores of the highly occurred healthcare terms which are graphically represented with the histogram of corpus keywords shown in Figure 2 and 3.

4.3 WordCloud or TagCloud

The top frequent healthcare terms are used to visualize the wordcloud in which the height of each term specifies the frequency of occurrence on the basis of TDM. The top most ailments of health-related tweets in wordcloud are HIV, malaria, aids, diarrhea, vomiting, coma, nausea and headache.

Table 6. Association between the ailments with 98% correlation

Function	findAssocs(dtms, c("sweating", "diarrhea"), corlimit=0.98)								
	$sweating								
	coma	Headache	Malaria	nausea	Vomiting	Across	Dies	Killer	reducing
	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
	$diarrhea								
Result	nausea	Vomiting	Coma	Dies	Headache		Malaria	Killer	reducing
	1.00	1.00	0.99	0.99	0.99		0.99	0.98	0.98

Table 7. Analyzing symptoms and ailments

	Corlimit	Ailments
Chest pain	0.89	HRDS (49%), HRAT (50%), CLG (4%), PNMa (9%)
Weakness	0.67	HRDS (28%), CLG (30%)
High blood pressure	0.96	HRDS (46%), CKD (9%), HIV (90%)
Nausea	0.93	HRAT (2%), CPT (9%), CKD (7%), DGu (7%), PNMa (2%), MLa (56%)
Abdominal pain	0.72	HRAT (0.6%), CPT (96%)
Fatigue	0.99	HRAT (10%), HIV (6%), DGu (4%), PNMa (8%)
Vomiting	0.38	DBS (15%), CPT (36%), CKD (2%), DGu (46%), PNMa (39%), MLa (42%)
Tiredness	0.59	DBS (5%), CLG (23%), CKD (58%)
Weight loss	0.76	DBS (60%), CPT (13%), AIDS (67%)
High fever	0.90	HIV (3%), DGu (6%), PNMa (2%), MLa (69%)
Diarrhea	0.89	AIDS (30%), PNMa (6%), MLa (28%)
Sweating	0.96	PNMa (46%), MLa (12%)

Heartdisease: HR_{DS}, Heartattack: HR_{AT} ,Lungcancer: C_{LG}, Pnemonia: PN_{Ma} ,Kidneycancer: C_{KD}, Pancreaticcancer: C_{PT}, Dengue: D_{Gu}, Malaria: M_{La}, Diabetes: D_{BS}

Table 8. Scoring Health keywords

Function	table (analysis$score)								
Range	-5	-4	-3	-2	-1	0	1	2	3
Scoring tweets	2	16	100	278	582	916	526	50	6

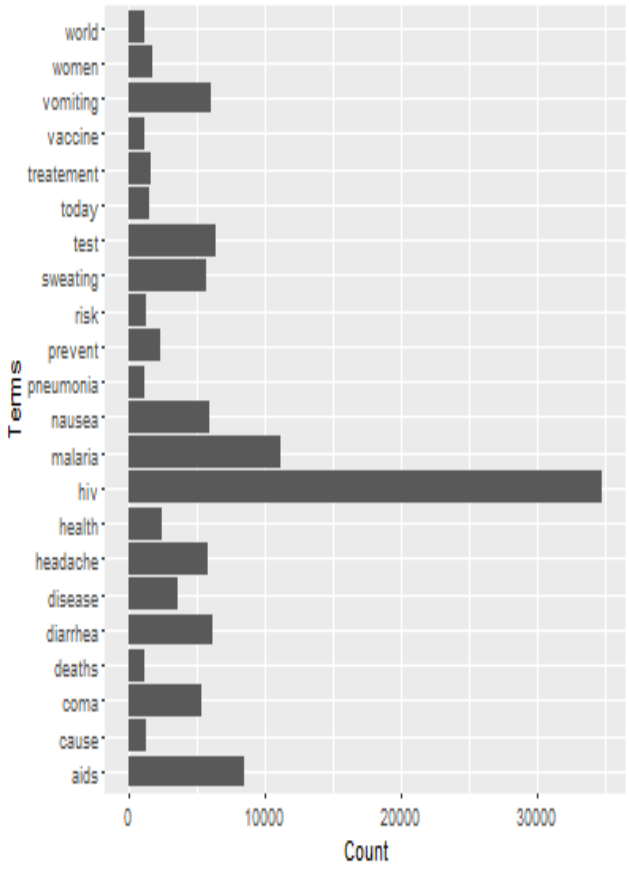


Figure 2. Health-related Term Frequency for common words and corresponding Plot formation as histogram

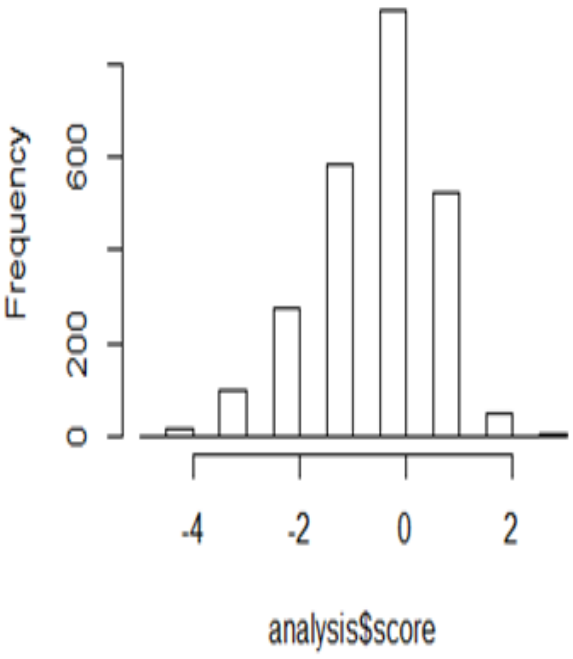
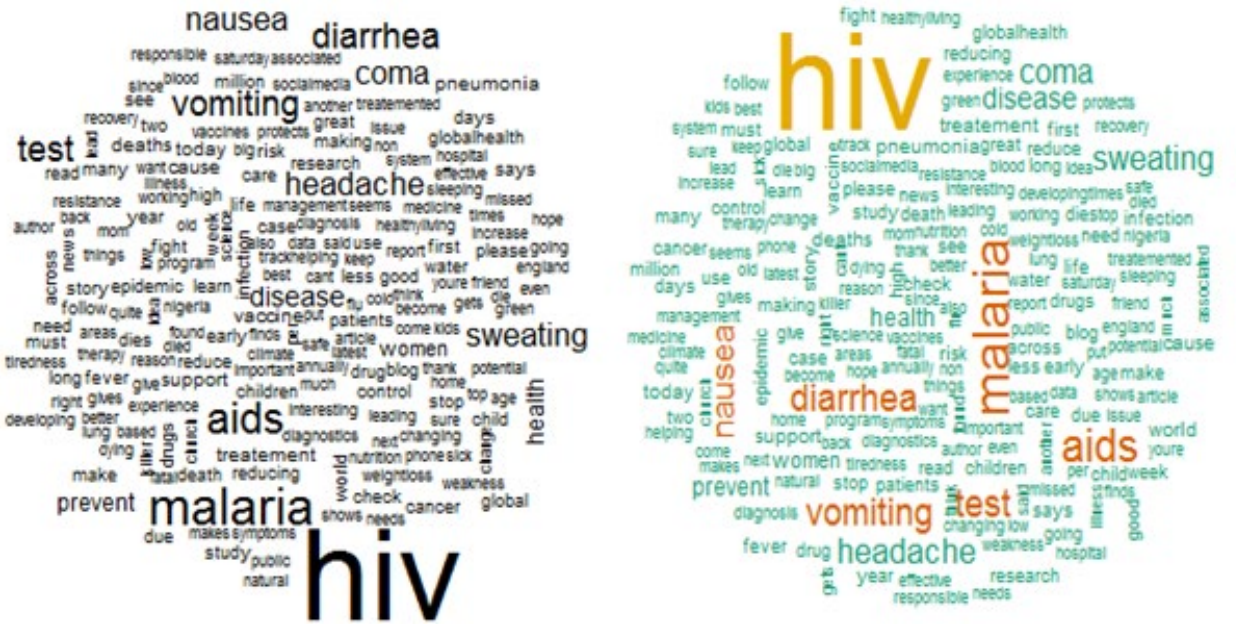


Figure 3. Histogram Scoring for Health related keywords used for analysis



(a) (b)

Figure 4. Commonality WordCloud of frequent occurred health terms

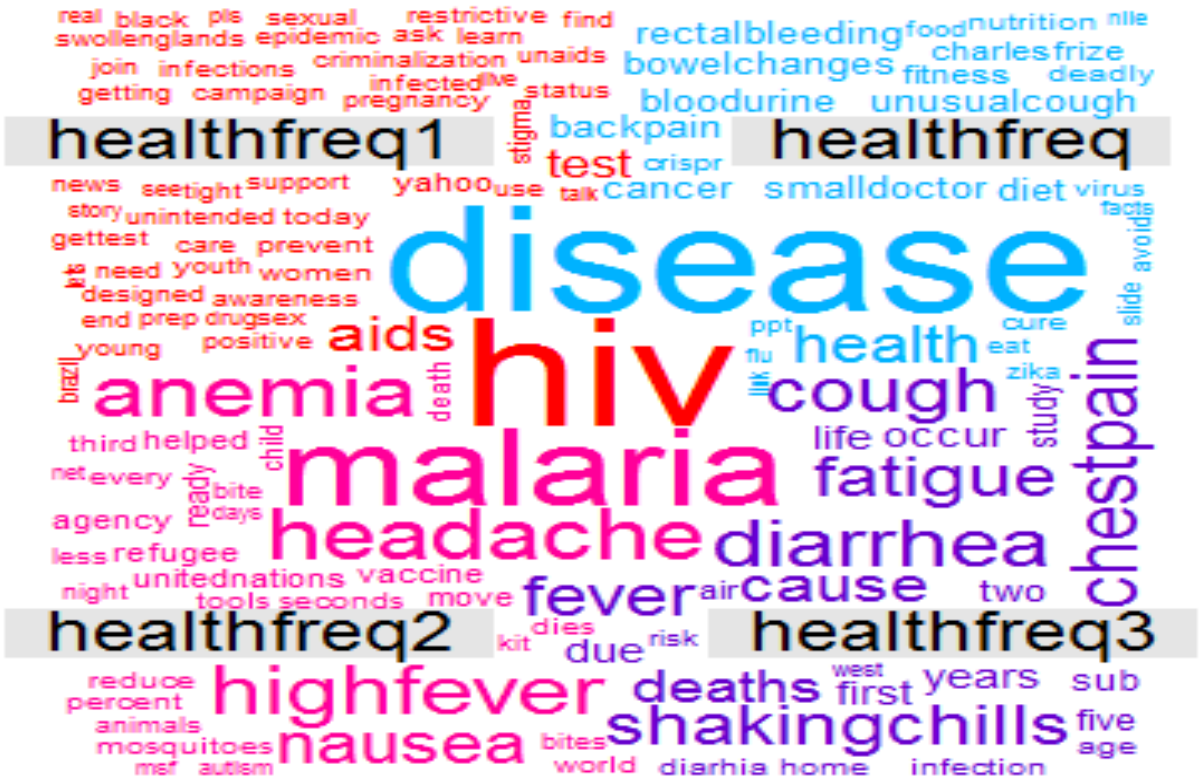


Figure 5. Comparing of health trends of WordCloud using proposed technique

As per [29] and [30] a technique by acquiring the Contrary Drug Antiphons and their fallouts has been presented. They also done the manual consideration of the drug related data and apply the opinion mining technique by integrating with the ML approach on NLP technique, which overcome the noisy data of contemporary systems. They used the ML technique to train and label the data set for build the classifier. They perform with classification algorithm used as classifier of SVM and NB. The Classifier performs with many of the attributes and dictionaries for achieving the result with measure score 00.794 as shown in Table 9.

Table 9. Measure score with classifiers

Attributes	SVM	NB
N1+Q2	00.658	00.658
N1+Q2+B3	000.676	00.661
N1+Q2+4S	00.675	00.704
N1+Q2+5M	00.734	00.676
N1+Q2+A6	00.632	00.676
N1+Q2+4S+5M	000.765	00.72
N1+Q2+M5+A6	00.779	00.691
N1+Q2+S4+M5+A6	00.794	00.706
M5+A6	00.571	00.644
Q2+M5+A6	00.691	00.661
N1+M5+A6	00.588	00.644
N1+Q2+B3+S4+M5+A6	00.691	00.66

Where N1: negatives, Q2: Que. Mark, B3: Score system for identify the Liu, S4: score WordNet,5: score MPQA, A6: score AFINN

In comparison to recent techniques as per citation [29, 30] and presented in table 9, the accuracy is 98% as compare to 79.4 % . It processes the raw information to bring out meaningful visions. The Figure 5 demonstrates the comparison of different terms over the WordCloud which concludes the result of correlated and highly occurred terms. The average accuracy in 98% as per table 10 through correlation Table 10. Association between the hiv ailments with 98% correlation

Table 10. Association between the hiv ailments with 98% correlation

Function	findAssocs (dtms, "hiv", corlimit=0.98)			
Result	$hiv			
	Aids	Care	Diagnosis	Epidemic
	1.00	1.00	1.00	1.00
	Patients	Prevent	Risk	test
	1.00	1.00	1.00	1.00
	Therapy	Control	Issue	news
	1.00	0.99	0.99	0.99
	Safe	Drug	Global	track
	0.99	0.98	0.98	0.98

5. Conclusions and Future Scope

This paper presents a framework for the real-time twitter feeds which are used to discover the highly occurring ailments and applies the association rules for finding the correlation between the Epidemic symptoms. An accuracy of 98% has achieved in presented framework while accuracy of maximum 79.4% was achieved in the existing frameworks. Some terms are most frequently used by people while posting the tweets on social media platform "Twitter". Visualization of such terms used to signify the frequently used terms are presented through "WordCloud" which provides the conjecture of the health tweets about the terms which are highly associated with the ailments and their symptoms. It processes the raw information to bring out meaningful visions. In this framework, the opinion mining has used to determine the tweets related to health issues.

In future, this research can be prolonged to analyse the sentiment phrases using a parser which will give the enhanced result to assess the sure, unsure, positive, negative and neutral sentiments because the similar terms may have dissimilar meaning in every sentence liable on it. Multiple similar phrases or synonyms words will also be handled to classify the tweets.

References

[1] Perrin, A. (2015). Social media usage. Pew research center, 52-68.

[2] Ji, X., Chun, S. A., Wei, Z., & Geller, J. (2015). Twitter sentiment classification for measuring public health concerns. Social Network Analysis and Mining, 5(1), 13.

[3] Kashyap, R., & Nahapetian, A. (2014, November). Tweet analysis for user health monitoring. In 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH) (pp. 348-351). IEEE..

[4] Zhou, X., Tao, X., Yong, J., & Yang, Z. (2013, June). Sentiment analysis on tweets for social events. In Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 557-562). IEEE.

[5] Dredze, M. (2012). How social media will change public health. IEEE Intelligent Systems, 27(4), 81-84.

[6] Tuan, T. M., Chuan, P. M., Ali, M., Ngan, T. T., & Mittal, M. (2018). Fuzzy and neutrosophic modeling for link prediction in social networks. Evolving Systems, 1-6.

[7] Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. Computers in human behavior, 54, 351-357.

[8] Khan M S A, Abdullah S, (2018).Interval-valued Pythagorean fuzzy GRA method for multiple attribute decision making with incomplete weight information ,International Journal of Intelligent Systems 33(8):1689-1716..

[9] Khan, M. S. A., Abdullah, S., Ali, A., Amin, F., & Hussain, F. (2019). Pythagorean hesitant fuzzy Choquet integral aggregation operators and their application to multi-attribute decision-making. Soft Computing, 23(1), 251-267.

[10] Kashyap, R., & Nahapetian, A. (2014, November). Tweet analysis for user health monitoring. In 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH) (pp. 348-351). IEEE.

[11] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.

[12] Kaur, P., Sharma, M., & Mittal, M. (2018). Big data and machine learning based secure healthcare framework. Procedia computer science, 132, 1049-1059.

[13] Ji, X., Chun, S. A., Wei, Z., & Geller, J. (2015). Twitter sentiment classification for measuring public health concerns. Social Network Analysis and Mining, 5(1), 13.

[14] Goyal, L. M., Mittal, M., & Sethi, J. K. (2016). Fuzzy model generation using Subtractive and Fuzzy C-Means clustering. CSI transactions on ICT, 4(2-4), 129-133.

[15] Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In Proceedings of the 14th international conference on World Wide Web (pp. 342-351). ACM.

[16] Duhan, N., & Mittal, M. (2017, December). Opinion mining using ontological spam detection. In 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS) (pp. 557-562). IEEE.

[17] Wu, L., Moh, T. S., & Khuri, N. (2015, October). Twitter opinion mining for adverse drug reactions. In 2015 IEEE International Conference on Big Data (Big Data) (pp. 1570-1574). IEEE.

[18] Batool, R., Khattak, A. M., Maqbool, J., & Lee, S. (2013, June). Precise tweet classification and sentiment analysis. In 2013 IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS) (pp. 461-466). IEEE.

[19] Sharma, M., Singh, G., & Singh, R. (2019). Design of GA and Ontology based NLP Frameworks for Online Opinion Mining. Recent Patents on Engineering, 13(2), 159-165.

[20] Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In Proceedings of the 2003 conference on Empirical methods in natural language processing (pp. 105-112).

[21] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12), 2009.

[22] Jasleen Kaur Sethi & Mamta Mittal (2019) "Ambient Air Quality Estimation using Supervised Learning Techniques", SIS, EAI, DOI: 10.4108/eai.13-7-2018.159406.

[23] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).

[24] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011, June). Sentiment analysis of twitter data. In Proceedings of the Workshop on Language in Social Media (LSM 2011) (pp. 30-38).

[25] Kashyap, R., & Nahapetian, A. (2014, November). Tweet analysis for user health monitoring. In 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH) (pp. 348-351). IEEE.

[26] Kaur, P., & Sharma, M. (2019). Diagnosis of Human Psychological Disorders using Supervised Learning and Nature-Inspired Computing Techniques: A Meta-Analysis. Journal of medical systems, 43(7), 204.

[27] Sharma, M., Sharma, S., & Singh, G. (2018). Performance Analysis of Statistical and Supervised Learning Techniques in Stock Data Mining. Data, 3(4), 54.

[28] Mittal, M., Goyal, L. M., Hemanth, D. J., & Sethi, J. K. (2019). Clustering approaches for high-dimensional databases: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3), e1300.

[29] http://web.njit.edu/~xj25/eosds_beta/files/news_stopwords.txt

[30] http://web.njit.edu/~xj25/eosds_beta/files/profanity_list.txt

[31] Sharma, M., Singh, G., & Singh, R. (2017). Stark assessment of lifestyle based human disorders using data mining based learning techniques. IRBM, 38(6), 305-324.

[32] Lin Y, Zhang J, Wang X, Zhou A (2012) An information theoretic approach to sentiment polarity classification In: Proceedings of the 2Nd Joint WICOW/AIRWeb Workshop on Web Quality, WebQuality '12, 35–40.. ACM, New York, NY, USA

[33] Cunha A.A.L., Costa M.C., Pacheco M.A.C. (2019) Sentiment Analysis of YouTube Video Comments Using Deep Neural Networks. In: Rutkowski L., Scherer R., Korytkowski M., Pedrycz W., Tadeusiewicz R., Zurada J. (eds) Artificial Intelligence and Soft Computing. ICAISC 2019. Lecture Notes in Computer Science, vol 11508. Springer, Cham

[34] Gupta C., Jain A., Joshi N. (2019) A Novel Approach to Feature Hierarchy in Aspect Based Sentiment Analysis Using OWA Operator. In: Krishna C., Dutta M., Kumar R. (eds) Proceedings of 2nd International Conference on Communication, Computing and Networking. Lecture Notes in Networks and Systems, vol 46. Springer, Singapore

[35] Pathak A.R., Pandey M., Rautaray S. (2020) Adaptive Model for Sentiment Analysis of Social Media Data Using Deep Learning. In: Gunjan V., Garcia Diaz V., Cardona M., Solanki V., Sunitha K. (eds) ICICCT 2019 – System Reliability, Quality Control, Safety, Maintenance and Management. ICICCT 2019. Springer, Singapore

[36] Khan A. et al. (2020) Sentiment Classification of User Reviews Using Supervised Learning Techniques with Comparative Opinion Mining Perspective. In: Arai K., Kapoor S. (eds) Advances in Computer Vision. CVC 2019. Advances in Intelligent Systems and Computing, vol 944. Springer, Cham