Indian Journal of Science and Technology, *Vol 11(4), DOI: 10.17485/ijst/2018/v11i4/119594,* January *2018*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Initiative to Identify Depression using Sentiment Analysis: A Machine Learning Approach

Akriti Sood[1], Madhurima Hooda[1], Saru Dhir[1] and Madhulika Bhatia[2]

[1]Department of IT, Amity University, Noida – 201313, Uttar Pradesh, India;
asood68@gmail.com, 10madhurima@gmail.com, sdhir@amity.edu
[2]FET, Manav Rachna International University, Faridabad - 121004, Haryana, India

Abstract

Objective: Online social and news media has turned into an extremely mainstream for clients to impart their insights. The objective of this paper is to propose a methodology through which sentiments can be analyzed. **Methods/Statistical analysis:** The sentiments are helpful for the identification of the depression. In this paper we proposed an algorithm through which tweets are extracted from twitter using R studio and then their sentiments are analyzed i.e. the scores are given to each sentiment by which we identify whether the person is depressed or not. This gives imperative data to basic leadership in different spaces. **Findings:** Sentiment analysis over Twitter offers associations and people a quick and powerful approach to screen the general population's sentiments towards them and their rivals. To evaluate the assumption examination over twitter we need a dataset that is been extracted from the twitter that would be publicly available for twitter sentiment analysis. We found that through twitter to extract tweets are scored based on their sentiments. The result is unique as we have proposed new algorithm through which twitter sentiments are scored. **Applications:** This sentiment analysis will be helpful to draw conclusion, whether the person is depressed or not. It can be helpful for prescreening test, diagnostic tool and automation monitoring system.

Keywords: Depression, Sentiment Analysis, Tweets, Twitter

1. Introduction

Depression is evaluated to impact 350 million people the world over. Depicted by assessments of outrageous pity or absence of concern, hopelessness baffles fundamental points of view of consistent everyday presence and can in like manner provoke suicide. The primary wellspring of disable, it is assessed that despairing expense generally $50 billion in lost benefit in the United States in 2010 alone.

Regardless of its inescapability, sadness is starting at now under diagnosed and undertreated. By a wide margin the majority of those resolved to have distress improve after proper treatment, in any case, not as much as half of those engaging with despairing get such treatment.

Twitter is extremely prominent person to person communication stages where a huge number of clients can give their perspectives. Long range informal communication benefits logically stretch out in more topographical districts and enter progressively in different portions of the populac[1]. The rate of data that is all things considered made on online stages increments. Correspondence and collaborations in online networking much of the time reflect genuine occasions. The user base of social networks gets larger and generating content about real-world events in real-time. Thus, social media streams become perfect sensors of real-world events. An enormous amount of available data requires information filtering for drilling down the relevant topics and events.

With the rise of online networking, the execution of sentiment analysis has turned out to be progressively basic[2]. In the present business rivalry, fashioners, designers, sellers and deals agents of new data items need to deliberately ponder whether and how do their items over upper hands[3].

**Author for correspondence*

Twitter, with more than 500 million enrolled clients and more than 400 million messages for each day[4], has turned into a gold dig for associations to screen their reputation and brands by extricating and investigating the slant of the tweets posted by the general population about them, their business sectors, and contenders[5]. Creating exact estimation investigation techniques requires the formation of assessment datasets that can be utilized to evaluate their exhibitions. Over the most recent couple of years a few assessment datasets for Twitter assumption investigation have been made freely accessible[6].

2. Literature Survey

Research has been done on sentiment analysis through social media and many algorithm has also been developed, to identify depression via texts and images.

Some of the researches done are as follows:

- Model to symbolize the tweets as goal, +ve and -ve. Classifier in view of the multinomial naive bayes approach that utilizations highlights like n-gram and pos-labels. The practise set they applied become much less productive because it contains just tweets having emojis[7]. They made a twitter corpus by way of gathering tweets utilizing twitter api and consequently explaining the ones tweets utilizing emojis. Utilising that corpus, they constructed up a supposition.
- A two-stage programmed sentiment analysis technique for arranging tweets. They organized tweets as target or subjective and afterward in second level, the subjective tweets were named effective or terrible. The detail space utilized covered retweets, hash tags, connection, accentuation and shout stamps along with highlights like earlier extremity of phrases and pos[8]`.
- Built up a 3-path show for grouping sentiment into positive, negative and impartial classes. They tried various things with models, for instance, unigram display, an element-based version and a tree component primarily based version. For tree component based totally version they spoke to tweets as a tree. They include based totally model uses one hundred highlights and the unigram display makes use of greater than 10,000 highlights. They touched base on an end that highlights which consolidate in advance extremity of phrases with their parts-of-speech(pos) labels are most essential and plays a noteworthy rolein the order mission. The tree part primarily based model outflanked the opposite two fashions[9].
- Applied twitter api to gather twitter statistics. Their preparation data falls in 3 specific classifications (camera, film, cellular). The statistics is marked as nice, negative and non-opinion. Tweets containing opinions had been separated. Unigram naive bayes show became achieved and the naive bayes streamlining autonomy suspicion changed into utilized. They moreover disposed of needless highlights by means of utilizing the mutual information and chi square detail extraction strategy. At lengthy remaining, the advent of a tweet is expected. I.e. Superb or poor[10].

3. Methodology

A qualitative analysis was conducted to explore depression on Twitter by using R studio[11]. R is additionally a programming dialect proposed for profound measurable examination. It is open source and accessible crosswise over various stages, e.g., Windows, Mac, Linux. It is presently utilized as a part of an assortment of uses including representations and information mining. You can utilize R to separate and envision Twitter information.

Steps to create an app to extract data from twitter:

Step1: you may need a twitter application and for this reason a twitter account. In case you don't have a twitter account, please sign up.

Step2: use your twitter login identification and password to register at twitter developers.

Step3: navigate to my programs in the top right-hand corner.

Step4: create a new application.

Step5: round out the new software body. Names must be unique, i.e., nobody else must have utilized this name for their twitter application. Supply a brief portrayal of the application. You can change this later if important. Input

your website online or blog address. Callback url can be left clean. Once you've got completed this, ensure you have perused the "designer guidelines of the street" ad spot, check the "sure, i concur" field, fill within the captcha and faucet the "make your twitter application" capture.

Step6: word the estimations of client key and patron mystery and keep them useful for some time later. You ought to hold these secrets. If each person in some way controlled to get these keys, they might competently get on your twitter account.

Step7: packages to be set up: r comes with a popular set of applications. Some of other programs are to be had for down load and set up. For this post, we are able to need the following packages:

- Roauth: provides an interface to the oauth 1.0 specification, allowing customers to authenticate via oauth to the server of their desire.
- Twitter: gives an interface to the twitter web api.

Step8: make and keep twitter proven qualification query. At the off threat which you are a home windows customer, you need to get "cacert.pem" report. Down load the "cacert.pem" record from the predetermined url and save it in your working registry. At that point make a query "cred" a good way to spare the showed protest for later classes and begin the handshake. That is the area you may enter the consumer key and consumer secret from the initial step. As soon as the handshake is completed it wills manual you to a link in the help window.

Step9: navigate to the required hyperlink to authorize app and click "authorize app".

Setup_twitter_oauth (consumerkey,consumersecret,accesstoken,accesst ok ensecret)

Note the pin quantity generated.

Step10: load "twitters authentication.rdata" record to your consultation and run register witteroauth. This must return "proper" demonstrating that all is first rate and we can continue. At that factor we set elements, one for the hunt string, which can be a hash tag or consumer say, and the second issue is the number of tweets we need to split for research[11]. Make use of search twitter to look twitter in mild of the provided searching for string and repair a rundown. The "lang" parameter is applied under to restriction tweets to the "English" dialect.

4. Proposed Algorithm

Following are the steps for the identification of sentiment analysis:

Step1: After extracting data from twitter use function write.csv(tweets.df, "tweets.csv"). The data will display in excel file.

Step2: my sentiment<get_nrc_sentiment (tweets) get_nrc_sentiment () will get the sentiments of that same tweets and the values will be store in the sentiment

Step3: sentiment scores<-cbind("sentiment"= row names(sentiment scores),sentiment scores) will arrange the sentiments and scores in rows and columns and store the scores in sentiment scores

Step4: the line sentiment scores will display the sentiment scores

Step5: str() function will give all the detailed information about the tweets like Who wrote the tweet and text of the tweet, favorite , count, when tweet created etc.

Procedure is shown in Figure 1, how tweets are extracted, and sentiments are analyzed.

5.Results

Figure 2 shows the sentiment analysis of the tweets. The datasets from the twitter is evaluated through the proposed algorithm and following result is shown above. In this sentiment have scored such as positive negative and natural for positive emotions you get score but if the emotion is negative such as anger, disgust etc. you don't attain scores.

Table 1 shows you what each emotion is scored, the tweets extracted are scored based on the sentiment. There are different scores for different emotions through which sentiments are analyzed.

Table 2 shows the date and time and id of the tweets. This is database through which the sentiments are analyzed.

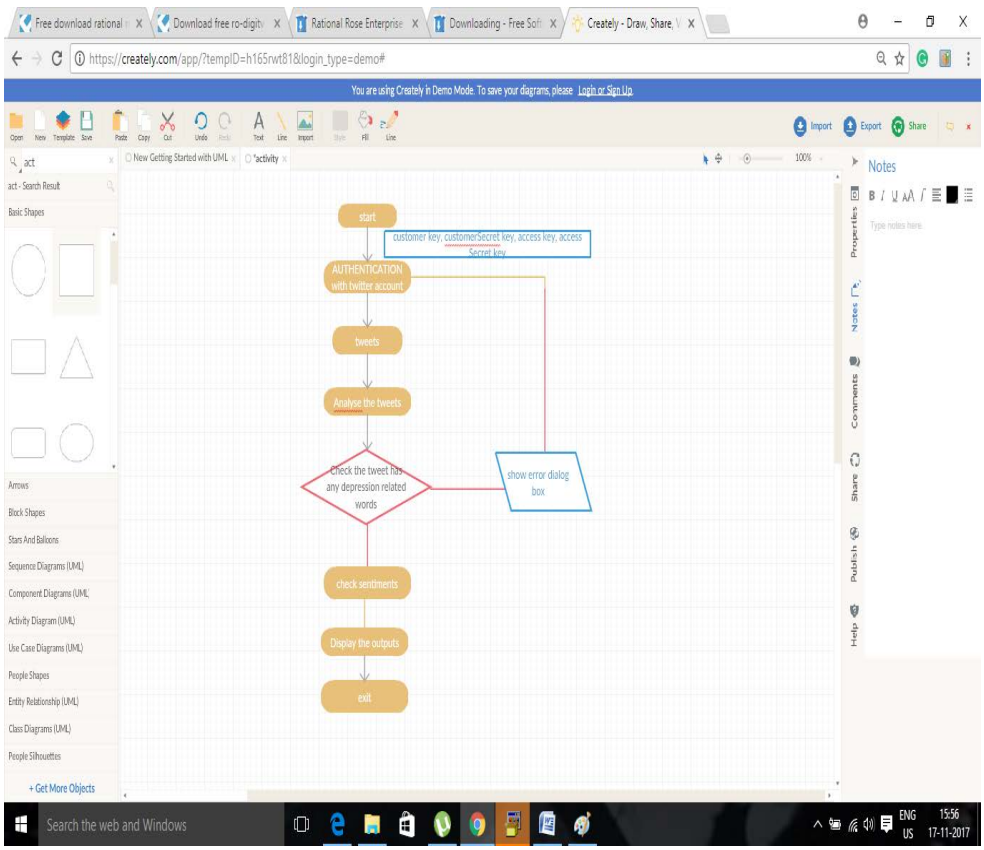

Figure 1. Process for Sentiment Analysis.

```
Console   Terminal
~/R/win-library/3.2/file3144a901904/
o í ½í¸,' in 'utf8towcs'
> mysentiment <- get_nrc_sentiment("i got a cute little beer this summerbut very safe to say that i was not sad to s
ee it")
> sentimentscores <- data.frame(colSums(mysentiment[,]))
> names(sentimentscores) <- "scores"
> sentimentscores <- cbind("sentiment" = rownames(sentimentscores),sentimentscores)
> rownames(sentimentscores) <- NULL;
> sentimentscores
      sentiment scores
1         anger      0
2  anticipation      0
3       disgust      0
4          fear      0
5           joy      2
6       sadness      0
7      surprise      0
8         trust      1
9      negative      0
10     positive      3
> str(tweets.df)
'data.frame':   150 obs. of  16 variables:
 $ text         : chr  "@nerushi_ Same im just really really sad" "I got a cute little beer gut this summer but it's
 very safe to say that I was NOT sad to see it go \xed<U+00A0><U+00BD>\xed<U+0"| __truncated__ "RT @FeshPinceBot: [s
ad piano music playing] Stupit, stupit, Uncle Fill, Uncle Fill..." "RT @jonmichaelolse1: Sick and sad. All the more
reason that we need to #banislam #BanSharia https://t.co/CCsWZcYCJU" ...
 $ favorited    : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ favoriteCount: num  0 0 0 0 0 0 0 0 0 0 ...
 $ replyToSN    : chr  "nerushi_" NA NA NA ...
 $ created      : POSIXct, format: "2017-11-07 18:58:25" "2017-11-07 18:58:25" "2017-11-07 18:58:24" "2017-11-07 18:
58:24" ...
 $ truncated    : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ replyToSID   : chr  "927973413314023429" NA NA NA ...
```

Figure 2. Results of Sentiment Analysis.

Table 1. Sentiment scores

Sentiment	Scores
Anger	0
Anticipation	0
Disgust	0
Fear	0
Sadness	0
Joy	2
Surprise	0
Trust	1
Negative	0
Positive	3

Table 2. Date, time, created and id of the tweets

Favourite	Fav Count	reply To SN	created	truncated	reply To SID id
1 FALSE	0	<NA>	2017-11-0718:58:25	FALSE	<NA 927973413314023429
2 FALSE	0	<NA>	2017-11-07 18:58:25	FALSE	<NA>927973530666401792
3 FALSE	0	<NA>	2017-11-07 18:58:24	FALSE	<NA>927973529383067648
4 FALSE	0	<NA>	2017-11-07 18:58:24	FALSE	<NA>927973528384794624
5 FALSE	0	<NA>	2017-11-07 18:58:24	FALSE	<NA>927973525381623808
6 FALSE	0	<NA>	2017-11-07 18:58:23	FALSE	<NA>927973525012582400

Table 3. Source and name of the Tweets

Reply To UID	status Source	screen Name
<NA>	Twitter for Android	arimasbootie
<NA>	Twitter for iPhone	LexyyNelson
<NA>	Twitter Web Client	WhereDaPuff
<NA>	Twitter for iPhone	RichardH1919
<NA>	Twitter for iPhone	elissabeebex
<NA>	Twitter for Android	SAD_DRAGON

Table 4. The count of the sentiments and retweets

Retweet	Count	is Retweet	retweeted	longitude	latitude
1	0	FALSE	FALSE	NA	NA
2	0	FALSE	FALSE	NA	NA
3	2	TRUE	FALSE	NA	NA
4	5	TRUE	FALSE	NA	NA
5	1	TRUE	FALSE	NA	NA
6	0	FALSE	FALSE	NA	NA

Each tweet is given a unique id. It also shows the whether the tweet is replied to which SID.

Table 3 shows replied tweets unique ID, their source and screen name. It shows the name of the person tweeting and source from where the tweet is received.

Table 4 shows the retweet and their count i.e. scores based on the sentiments.

6. Conclusions

In our research, we have proposed an algorithm which can differentiate between depressed and non-depressed person through social media i.e. twitter by their status and messages. It is really helpful in real world for trauma centers as prescreening of the person can be done. People nowadays like to upload their day to day activities in social media, by this we can identify people suffering from depression and

help them to lead a better and prosperous life, as mental illness can be harmful for the society and oneself.

7. Future Work

In our research we have developed an algorithm, but many more work can be done, in this field, some of them are:

- We will develop an application through which depression analysis can be done and a direct message will be delivered to their guardians.
- We have done research using text message, analysis can be done through photos and voice also.
- Research can be done in other social media sites also.

8. Key Recommendation

Algorithm proposed on detection of depression through twitter can be helpful in many ways to mankind.

Some of the fields or examples where this research can be helpful are:-

- **Pre-screening quiz**[12]- It can be helpful as people generally take up quiz on social media like "Who they will marry" etc. it can be monthly quiz to examine the mental health of the person. It can be conducted online in schools, colleges, work places etc.
- **An extra diagnostic tool**[12] - It can be helpful for doctors, as in early stages it is difficult to determine the depression, so this algorithm can be used for early detection of depression.
- **Automatic monitoring of system**[12] - It can not only be helpful as an early detection of depression but also monitor the status of his/her health.

9. Acknowledgement

Authors are grateful to Amity University and staff for their help.

10. References

1. Paparrizos J, White RW, Horvitz E. Screening for Pancreatic Aden carcinoma Using Signals from Web Search Logs: Feasibility Study and Results. Journal of Oncology Practice. 2016; p. 1-6.
2. Visualization of Egyptian Revolution on Twitter. Date accessed: 15/02/2011: Available from: crossref
3. Wall M, Zahed SEL. I'll be waiting for you guys: A You Tube call to action in the Egyptian revolution. Journal of Communication. 2011; 5:1333-43.
4. Sakaki T, Okazaki M, Matsuo Y. Earthquake shakes twitter users: Real-time event detection by social sensors. In Proceeding WWW: 19th ACM Association for Computing Machinery International Conference World Wide Web. 2010; p. 851-60.
5. Conover MD, Gonçalves B, Ratkiewicz J, Flammini A, Menczer F. USA: MA: Boston: Predicting the political alignment of twitter users. IEEE International Conference on Social Computing. 2011 October, p. 1-8. crossref
6. Saif H, Fernandez M, Yulanand H, Alani H. Evaluation Datasets for Twitter Sentiment Analysis A survey and a new dataset, the STS-Gold. 1st International Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI. 2013; p. 1-13.
7. Pak A, Paroubek P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh Conference on International Language Resources and Evaluation. 2010; p. 1-7.
8. Barbosa L, Feng J. Robust Sentiment Detection on Twitter from Biased and Noisy Data. COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters. 2010; p. 36-44.
9. Agarwal A, Xie B, Vovsha O, Rambow R, Passonneau. Sentiment Analysis of Twitter Data. In Proceedings of the ACL Workshop on Languages in Social Media. 2011; 139(11):30-38. PMid:21552405 PMCid:PMC3083532
10. Po-Wei L, Bi-Ru D. Opinion Mining on Social Media Data. IEEE 14th International Conference on Mobile Data Management. 2013; 2:91-96.
11. Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. PLoS ONE. 2010; 5(9):12948. crossref PMid:20856792 PMCid:PMC2939797
12. Tang E. Identifying Signs of Depression on Twitter. Class of Dobin Prize Submission. 2016; p. 1-8.