

# Sentiment Analysis on Twitter Data using R

Sonia Saini  
Amity Institute of Information  
Technology, Sector-125,  
Noida, U.P, India  
ssani2@amity.edu

Ritu Punhani  
Amity School of Engineering &  
Technology, Sector-125  
Noida, U.P, India  
rpunhani@amity.edu

Ruchika Bathla  
Amity Institute of Information  
Technology, Sector-125  
Noida, U.P, India  
rbathla@amity.edu

Vinod Kumar Shukla  
Department of Engineering  
and Architecture, Amity University  
Dubai, UAE  
vshukla@amityuniversity.ae

**Abstract**—Nowadays social networking sites are at the blast from where huge amount of information is produced or retrieved. 90% people of the world are sharing their perspectives every day on micro blogging sites, since it contains short and simple expressions. The various devices, mobiles, laptops, tabs and other IoT data gadgets generate huge volume of data and Microservices based web applications running on these have made it simpler for us to get any kind of data at any time and from any place. Social media is also used for expressing our opinions for the products and services. The feedbacks and ratings of millions of the social site users can be collated to extract their attitudes and sentiment towards any products or services and use that information for future market and business improvement or domain analysis. Mining user's opinion from social media is a difficult task; it can be refined into numerous ways. In this paper, an open source approach is presented which we have collected tweets from Twitter API and then pre-processed, analyzed and visualized these tweets using R. To analyze sentiments of tweets we are utilizing a statistical tool, R programming. This sentiment analysis is based on text data retrieval from streamed web and then classifying people perspectives in eight distinct classifications of feeling (disgust, fear, anger, anticipation, sadness, trust, surprise) and two unique sentiments (positive and negative).

**Keywords**—Sentiment Analysis, E-Healthcare, R Programming, Twitter

## I. INTRODUCTION

These days, technology has got its new and higher pace. This development has changed human's way of expressing their opinions, sentiments and views and the platforms in which they do so [1]. We know that there are almost around 111 micro blogging sites [10].

Micro blogging websites are just social networking webpage on which people write regular and short posts. One of the most famous micro blogging services is twitter where people can post and read messages which can be 148 characters long [4]. Messages in twitter are known as Tweets which we will utilize as crude information. A strategy that we will use automatically extracts tweets into neutral, negative, and positive sentiments [10]. By utilizing the sentiment analyses the user can able to know the criticism about the services or item before buying it and the firm can know about the feeling of clients about their items, with the goal that they can analyze consumer satisfaction and according to that they can improvise their items. Today around approx 6500 tweets are tweeted every

second, which roughly brings out 561.6 million tweets for every day. These streams of tweets are mainly noisy reflecting multi topic, changing states of mind information in unfiltered and unstructured format. Analyzing unstructured data is in itself a difficult task and extracting useful information from it's a big challenge. For doing this, there is a need of powerful tools and technologies which can help to handle millions of tweets and extracting sentiment from them. There are many different possible ways to do this. In this paper we are using *R language* to do sentiment analysis [11]. R is an open source approach used for analyzing on-line reviews to perform sentiment analysis and text mining [4].

Sentimental Analysis is a strategy to explore whether a gathered content is in positive, negative or neutral state. Essentially, it involves examining the emotions related with a piece of writing for any topic. Sentiment analysis is used to check the opinions, taste, views and interest of individuals by seeing diverse prospective, for example, celebrity, politicians, foods, places, or some other topic [5]. In sentimental analysis we usually classify everyone's mood in various classifications.

Distinct levels Sentiment Analysis can be applied are as follows:

- *Level 1.*  
*Sentence level:* It recognizes neutral, negative and positive sentiment for every line [8].
- *Level 2.*  
*Document level:* It recognizes the entire record of sentiment as one entity or one unit neutral or negative or positive [8].
- *Level 3.*  
*Aspect level:* It is utilized in case of the availability of traits inside post, input text or entity. Each trait can hold a sentiment in its own. It can prompt a superior analysis and results if taken into consideration. Some sentiment analyses techniques are applied for grouping on this level where all attributes having a similar sentiment outcome are gathered together [8].
- *Level 4.*  
*User level:* It handles the social connections between various clients by utilizing graph theory [8].

## A. Need of Sentiment Analysis

Sentiment analysis is increasingly very important because of rise of *web-based life*. Sentiment analysis can be utilized in all sorts of task and strategies. Some of them are in the field of:

- *Business*: In marketing field organizations utilize it to build up their strategies, to understand user's sentiments towards their items or brand, how everyone react to their campaigns or item launches and why shoppers don't buy some items.
- *Politics*: In political field, it is utilized to monitor political view, to recognize inconsistency and consistency between activities and statements at the administration level. It can be utilized to predict election results as well!
- *Public Actions*: Sentiment analysis is also utilized to monitor and analyze social wonders, for the spotting of situations which are possibly dangerous and deciding the general state of mind of the bloggers.

## II. RELATED WORK

This section illustrates other similar work done related to sentiment analysis.

In this paper [2] author describes the importance and applications of opinion mining and sentiment analysis in social networks and the basic concepts, challenges and comprehensive study in different sections.

In [6] author describes the preprocessing steps which have to be applied to extract bags of words from Twitter data in detail and propose a topic-based sentiment analysis approach. The paper has focused on exploiting the results of the default parameter for the topic modeling method.

In another paper author presents an algorithm to convert "bulk of data" available from social media (Twitter) into useful data and extract information by processing it to suit our requirement. Other benefits related with the automatic sentiment analysis presented, include subjects who express their opinions frequently have much distinct opinions than others. All thoughts are extracted in real-time, letting for earlier response times to market changes and for full time-based data because of which it become possible to plot trends over time using R language on twitter. The obtained analysis can be used to infer population attitudes to generalize the prevailing trends of the market and make predictions regarding profit making sectors [7].

In [9] authors describe the main objective of this paper was to describe and design system for twitter data analysis and visualization by using R and the big data processing technologies called Hadoop. They developed a set of analytical representation which helps user to identify product, people, services and movies data and can gain insights from it and they also took a set of visualizations,

implemented in Shiny web applications which helps to integrate user interface with RHadoop.

In [3] authors describe the utilization of sentiment analysis methods on text-based information that is related to health care. This information is ideally extricated from web sources. The sentiment analysis for health care identifies the areas that are appreciated, criticized, proposed with improvements or reasoned upon execution.

In another [12] author studied the sentiment analysis on mobile reviews.

## III. DATA COLLECTION AND CLEANING

Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes. Accurate data collection is essential in maintaining the integrity of research.

Data cleaning refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. In this paper we fetched 3000 tweets from twitter for analyzing sentiments of twitter user's for E Healthcare and some taken diseases (Heart Attack, Depression, Diabetes, Cancer, Tuberculosis) by using R. Then we have converted those collected tweets into data frame and then perform data cleaning on it because there is so much useless data in it which are not going to be used during sentiment analysis like semi comas, blank spacing and so on. Data collection and data cleaning is the very important or essential part of sentiment analysis. Without data collection sentiment analysis is impossible and without data cleaning sentiment analysis is useless because that will be inaccurate.

### A. Tools and Packages Used

The project is done in RStudio GUI using various packages as listed below:

1. *twitteR*: It is an R package which provides access to the Twitter API. Most usefulness of the API is supported, with an inclination towards API calls that are more valuable in data analysis instead of day by day interaction.
2. *ROAuth*: It provides an interface to the OAuth 1.0 specification enabling clients to authenticate via OAuth to their preferred server.
3. *plyr*: It is a set of tools for a common set of issues: we have to part up an information structure which is big into homogeneous pieces, apply a function to each piece and after that combine every one of the outcomes back together.
4. *dplyr*: A consistent, quick tool for working with data frame like both out of memory and in memory, objects.





that positive score is 477 and negative score is 304. Fig. 8 shows that positive score is 303 and negative score is 275.

Similarly, results were observed for positive score of 438 and negative score 326 in diabetes. In case of Tuberculosis positive score of 289 and negative score of 254 was observed. In case of cancer, positive score is 327 and negative score is 251. In case of Depression- positive score is 361 and negative score is 380.

TABLE I: POSITIVE & NEGATIVE SCORE OF DISEASES

Diseases	Positive score	Negative Score
E healthcare	477	304
Heart Attack	303	275
Tuberculosis	289	254
Diabetes	438	326
Cancer	327	251
Depression	361	380

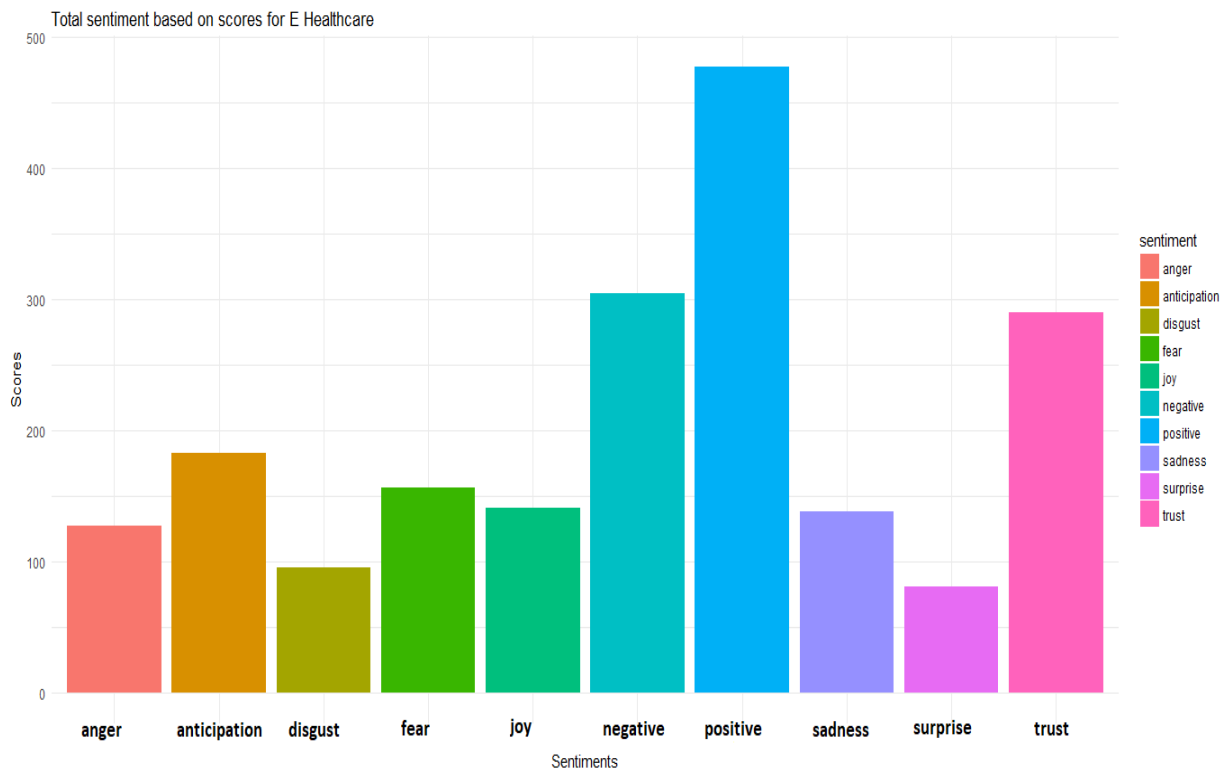


Fig. 7 Histogram of E Healthcare

## V. CONCLUSION AND FUTURE WORK

By the above given sentiment analysis, we conclude that for E-healthcare people have positive opinion and sentiment as its positive sentiment score (477) is higher than negative sentiment score (304).

This paper takes the perspective of analyzing social sentiment of various ailments. This is a very curated set and does not account for location aggregation and though it shows a varied range of 10 different sentiments, there is still scope further for a large-scale analysis along with weightage assignment to the different sentiments for a more refined analysis as future work for this research.

Location aggregation-based analysis will provide exact insight about region specific sentiments, while weightage assignment will provide clear segregation of sentiment boundaries.

This paper research work can also be further extended to analyze if multiple sentiments in one document like “love” and “like” are synonymous to the sentiment “positive” and as such multi-label classification machine learning can be done as part of further research.

## REFERENCES

- [1] Federico Neri, Carlo Aliprandi, Federico Capeci, Montserrat Cuadros, Tomas By, “Sentiment Analysis on Social Media “, Published at ACM International Conference on Advances in Social Networks Analysis and Mining, 2012.
- [2] Sudipta Roy, Sourish Dhar, Arnab Paul, Saprativa Bhattacharjee, Anirban Das, Deepjyoti Choudhury,” Current Trends Of Opinion Mining And Sentiment Analysis In Social Networks”, International Journal of Research in Engineering and Technology, Volume 2, Special Issue 2, December 2013.
- [3] M. Taimoor Khan, Shehzad Khalid, “Sentiment Analysis for Health Care”, International Journal of Privacy and Health Information Management, 2015.
- [4] Eman M.G. Younis, “Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study”, International Journal of Computer Applications, Volume 112 – No. 5, February 2015.
- [5] Akshi Kumar and Teeja Mary Sebastian, “Sentiment Analysis on Twitter”, International Journal of Computer Science Issues, Volume 9, Issue 4, No. 3, July 2012.
- [6] Pierre Ficamos, Yan Liu, “A Topic based Approach for Sentiment Analysis on Twitter Data”, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 12, 2016.
- [7] Pooja Khanna, Sachin Kumar, Sumita Mishra, Anant Sinha, “Sentiment analysis: An approach to opinion mining from twitter

- data using R”, International Journal of Advanced Research in Computer Science, Volume 8, No. 8, 2017.
- [8] Kiruthika M., Sanjana Woonna, Priyanka Giri, “Sentiment Analysis of Twitter Data”, International Journal of Innovations in Engineering and Technology, Volume 6, Issue 4, April 2016.
- [9] Shubham S. Deshmukh, Harshal Joshi, Pranali Pandhare, Aniket More, Prof. Aniket M. Junghare, “Twitter Data Analysis using R”, International Journal of Science, Engineering and Technology Research, Volume 6, Issue 4, April 2017.
- [10] Varsha Sahayak, Vijaya Shete, Apashabi Pathan, “Sentiment Analysis on Twitter Data”, International Journal of Innovative Research in Advanced Engineering, Issue 1, Volume 2, January 2015.
- [11] Sonal Singh, Shyam S Choudhary, “Social Media Analysis: Sentiment Analysis Twitter Using R Language”, International Journal of Advances in Electronics and Computer Science, Volume 4, Issue 11, November 2017.
- [12] Onam Bharti, Mrs. Monika Malhotra, “Sentiment Analysis”, International Journal of Computer Science and Mobile Computing, Volume 5, Issue. 6, pages 625 – 633, June 2016.