Pulling social media twitter data into R

Roshan R. Karwa   Vishal V. Rathi   Parag P. Kadu   Ankush R. Deshmukh

Abstract- **R is software as well as language considered as a dialect of the S language produced by the AT&T BellLaboratories. R is a well- simple and useful programming language which includes conditional statements, loopslike repeat, while, for; user defined functions and input output facilities. R has a valuable data handling and storagefacility. R provides a set of operators like Arithmetical, Relational, Logical which can be used to perform operationson Data Objects like Vector, Factor, Array, List and Data frame. R provides a great, rational and included group oftools for data analysis. R provides graphical facilities for data analysis and demonstrates either directly at thecomputer or printing at the papers. This paper is about how to get setup to pull social media twitter data into R sothat we can do text analytics with it.**

Keywords-**Twitter, Data Extraction, Analytics, Social Media**

I. INTRODUCTION

Data is factual thing or entity. When data isorganized, processed or presented in a given context,it is called as Information. For example, Student isData whereas his/her percentage is Information asPercentage will be processed. We reside in a world that's drowning in data. Data is not only in Textformat but also in other formats like Image, Video,and Geographical Location etc. That is, Data isavailable in Structured as well as in unstructured dataand study of Data is called as Data Science. In otherterminology, Data science is the multidisciplinaryfield that focuses on finding actionable information inlarge, raw or structured data sets to identify patternsand uncover other insights. The data science domainlooks for to find out answers for areas that areunidentified and unforeseen. Analysis of this data isvery important. To analyze data, one need toappropriate platform where data will be stored anddata will be retrieved.On Social media like Facebook, Twitter, Instagramand Whatsapp, users express their opinions by meansof Posts, Status, Tweets, Stories or messages. It isvery important to analyze this data and first step inanalyzing data is Extracting Data from SocialMedia. This paper deals with extraction of data(tweets) of Twitter media using R Programminglanguage. [13]R which is both software and programminglanguage is provides different packages that helps toachieve objective. R is Interpreted language that is itis not a Compiled language, there is no need tocompile the code written, it will be directly run onconsole.

II. PLATFORM FOR PULLING DATA

RStudio is a open source integrated developmentenvironment (IDE) for R programming language. Itwas developed by JJ Allaire and Chief Scientist atRStudio is Hadley Wickham. It is accessible in twoeditions:
1) RStudio Desktop: The program is run locallyas a regular desktop application
2) RStudio Server: It allows accessing RStudiousing a web browser while it is running on aremote server.

III. LANGUAGE USED FOR PULLING DATA

Language used for pulling data from Social MediaTwitter is R Programming [1]. R is InterpretedProgramming language. Once you type commands, itwill be executed directly without compiling. One canperform many operations on R using R like Obtainingsatellite image of Particular area, Reading from file,writing into File, Analyzing data by generatingGraphs & Charts etc. This all functions can beperformed using appropriate package.R Packages contains collection of Functions and itssample data. These are accumulating under directorycalled "Library" in R environment.

To install Package, the syntax is

install.packages ("package_name")

The above command gets the respective package fromCRAN Website and then install in your Renvironment. While installing, it gives message toselect nearest location. Select India (https) if you arein India or otherwise.

IV. PACKAGES REQUIRED FOR PULLING DATA

There are 5 Packages required to do the required task,they are as follows: [8]
    1) twiiteR
    2) bitops
    3) ROAuth
    4) RCurl
    5) RJSONIO

1) **twitteR**: twitteR is provides right of entry to theTwitter API. Through this Package, Tweets can beextracted as per our

requirement, also in differentlanguage like English, Hindi etc. This task is intendedfor Data Analysis Purpose. There are also some otherfunctions like getting information of Followers, Following, timeline etc. [2]

2) **bitops:** It provides Functions for bitwise operationson vectors. Operations are such as Bitwise and,Bitwise Flip, Bitwise shift (left to right or right to left)etc. [3]

3) **ROAuth:** It provides an interface to the OAuth,OAuth token is required when the condition is of R tocommunicate with Online services. [4]

4) **RCurl:** When there is need of communication withonline services, it requires HTTP requests and thispackage provides functions to allow one to composegeneral HTTP requests. It also helps to do functionlike to fetch URIs, get & post forms, etc. It alsoprocess the output resulted by Web Server. It alsosupports redirects, cookies & other authentications.[5]

5) **RJSONIO:** This package allows conversion to andfrom data in JavaScript object Notation i.e. JSONformat which is requirement of some applications todeploy or to develop. [6][7]

V. TWITTER API

Twitter is what is happening now. Twitter"s developerplatform offers many APIs, tools, and resources thatenable you to harness the power of Twitter's open,global and real-time communication network. [9,10]To fetch tweets from twitter, one needs to have fourkeys:

　　　1. Consumer API Key
　　　2. Consumer API Secret Key
　　　3. Access Token Key
　　　4. Access Token Secret Key

To have above keys, one needs to create a twitter appon https://apps.twitter.com
Every application created on twitter API has uniqueabove keys. Consumer API Key and Consumer APISecret are used to authenticate requests to the TwitterPlatform. Always maintain the "Consumer API SecretKey" as a secret. This key should never be humanreadablein application. Access token key are used tomake API requests on own account's behalf. Nevershare Consumer API Key, Consumer API Secret

key,Token Access key & token access secret key withanyone. [12][13]

VI. IMPLEMENTATION DETAIL

Step1: Sign Up on Twitter website on tohttp://twitter.com/signup to create account on twitter.
Enter correct email id and mobile number whileregistration; also verify the same for further use. Aftersign up, Sign in with correct credentials i.e. User idand Password

Step 2: After creating account, go tohttps://apps.twitter.com to create app. Keep sign onwith your twitter account.

Step 3: After sign in at twitter API, apply fordevelopers account. Fill the information correctly,provide reason for which purpose you need twitterAPI, twitter keys. Successfully submit the same. Aftersubmission, the application status will be underreview. Twitter team will review application, and willthen approve after reviewing process. Then only, onecan able to create new app.

Step 4: If application is approved, Hit on the "CreateNew App" icon. Submit the details of personal use.After successful creation, following keys willgenerate

　　　1. Consumer API Key
　　　2. Consumer API Secret Key
　　　3. Access Token Key
　　　4. Access Token Secret Key

Keys description is given in Section 5 i.e. of TwitterAPI. Now all settings are done on Twitter Side.

Step 5: Open R Software SessionCall required packages in R environment i.e. import

　　　library in R Session.
　　　library(twitteR)
　　　library(bitops)
　　　library(RJSONIO)
　　　library(ROAuth)
　　　library(RCurl)

Use keys generated in Step 4 and store it in respectivevariables.

　　　ck<- "your_consumer_key"
　　　cs<- "your_consumer_secret"
　　　at<- "your_access_token"
　　　as<- "your_access_secret"

Setting authentication is very important and it is doneby below function.
setup_twitter_oauth(ck,cs,at,as)

Then analysis the data by our required use by applyingfunctions of twitter package.
To extract tweet, function searchTwitter will be usedin following way.
t <- searchTwitter("Narendra Modi", n=10 , lang='en')

This function will store Tweets which features"Narendra Modi" in „t‟ object. n=10 specifies 10tweets will be stored and lang= „en‟ specifies thelanguage of tweet will be English.

Screenshot of result obtained is follows:
Figure 1: Screenshot of Output (Extracting tweets based on **Narendra Modi** in R Environment)

VII. CONCLUSION

Today data is generating everywhere and by everyone,that also in great mass, it is very necessary to analyse data. Data is also of any format that is it is not instructured format. To analyze this big amount of data,R Studio is good platform and first step in analyse data is "Pulling social media twitter data into R".

VII. FUTURE SCOPE

After extracting data (tweets) from social media (twitter), in future it will be used to
1. Amalgamate Tweets into app or website?
2. Append the additional function to post Tweets from app or website?
3. Target an audience and advertise on the Twitter network?
4. Search for real time tweets of own choice
5. Find out Sentiments of Users

REFERENCES

1. R Software Manual https://cran.r-project.org/

2. Package „twitteR‟ by Jeff Gentry https://cran.rproject.org/web/packages/twitteR/twitteR.pdf

3. Package „bitops‟ S original by Steve Dutky <sdutky@terpalum.umd.edu> initial R port and extensions by Martin Maechler; revised and modified by Steve Dutkyhttps://cran.rproject.org/web/packages/bitops/bitops.pdf

4. Package „ROAuth‟ by Jeff Gentry https://cran.rproject.org/web/packages/ROAuth/ROAuth.pdf

5. Package „RCurl‟ by Duncan Temple Lang and the CRAN team https://cran.rproject.org/web/packages/RCurl/RCurl.pdf

6. Package „RJSONIO‟ by Duncan Temple Lang https://cran.rproject.org/web/packages/RJSONIO/RJSONIO.pdf

7. Sonal Singh, ShyamChoudhary, "SOCIAL MEDIA DATA ANALYSIS: TWITTER SENTIMENTAL ANALYSIS USING R LANGUAGE" in International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-4, Issue-11, Nov.-2017

8. ShubhamDeshmukh, Harshal Joshi, Pranali Pandhare, Aniket More, Prof.Aniket M. Junghare, "Twitter Data Analysis using R" inInternational Journal of Science , Engineeringand Technology Research (IJSETR) Volume6, Issue 4,April2017,ISSN:2278-7798

9. https://dev.twitter.com/docs

10. https://dev.twitter.com/overview/api.

11. VaddadiVasudha Rani1 and K. Sandhya Rani, "Twitter Streaming and Analysis through R" Indian Journal of Science and Technology, Vol 9(45), DOI:10.17485/ijst/2016/v9i45/97914, December 2016.

12. ShilpyGandharv, VivekRicchariya, Vineet Ricchariya, "Real Time Text Mining on

Twitter Data" in IJCAhttps://www.ijcaonline.org/archives/volume178/number3/gandharv-2017-ijca-915779.pdf

13. https://www.credera.com/blog/businessintelligence/twitter-analytics-using-r-part-1-extract-tweets/