Enhanced Social Media Metrics Analyzer Using Twitter Corpus as an Example

K.Jayamalini, M.Ponnavaikko

Abstract*: Social media is the collection of online communications channels dedicated to community-based input, interaction, content-sharing and collaboration. Websites and mobile applications dedicated to forums, microblogging, social networking, social bookmarking, social curation, and wikis are most popular and different types of social media. Social media become an essential part of human life. In business, social media is used to market products, promote brands, and connect to current customers and foster new business.Online social media is ubiquitous in nature. It allows people to use short text messages to express their opinions and sentiments about products, events and other people. For example, Twitter is an online news and social networking service where users post and interact with short messages, called "tweets". Therefore, nowadays social media become a potential source for business to find people's sentiments and opinions about a particular event or product. Social media analytics is the practice of gathering huge amount of digital data generated online from blogs and social media websites and analyzing them to find the insights and make business decisions. This paper focuses on development of enhanced social media metrics analyser using various latest methods and algorithms with the help of R language and R tool.

***Index Terms*: Social Media data, Opinion Mining(OM), Sentiment Analysis (SA), Metrics Analyser, Twitter.**

I. INTRODUCTION

Social media are computer-mediated [1] technologies that facilitate the creation and sharing of information, ideas, career interests and other forms of expression via virtual communities and networks. The variety of stand-alone and built-in social media services currently available introduces challenges of definition; however, there are some common features:

User-generated content, such as text messages, multimedia contents such as digital photos or videos, and data generated through all online interactions, are the lifeblood of social media[1][2]. Users create service-specific profiles for the website or mobile app that are designed and maintained by the social media organization. Social media facilitate the development of online social networks by connecting a user's with other individuals or groups.[1][3]

Here are some prominent examples of social media:

- Facebook - popular free social networking website or mobile app that permits registered users to create their profiles, upload multimedia contents, send messages and communicate with friends, family members and colleagues.
- Twitter - a free microblogging web service that permits registered users to broadcast short messages called tweets about a particular event, product, service or a person.
- Wikipedia - a free, open source online encyclopedia created by Wikipedians, a community of users, through their collective effort. Registered users can create an artifact for publication, but anyone can able to edit the articles.
- LinkedIn – professional online social networking site specially for the business community.
- Instagram - social networking site or app for sharing photos and videos.
- Youtube – video sharing website

A.Every Minute Activityof Social Media

The '60 seconds'[2] info graphic in figure below visualizes just one minute activities on the web. The number of Google searches, Facebook posts and messages sent in WhatsApp in a minute is truly unbelievable! Analysis of the huge amount of digital data generated by social media help in taking business decisions.

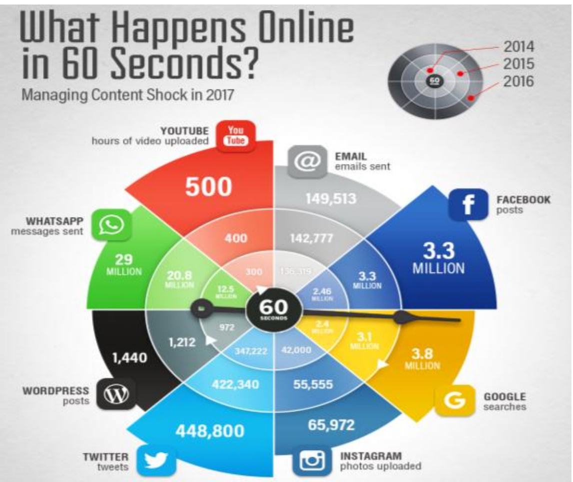


Fig.1. Social Media Activity in every minute

II. DIFFERENT TYPES OF SOCIAL MEDIA ANALYTICAL TOOLS

Social Media is important for business. Social media helps business to

a) create successful social campaigns using marketing analytics
b) recognise influencers for their brand, product, service & industry
c) compare key performance metrics and to find strengths, weaknesses of

Revised Manuscript Received on May 10, 2019

K.Jayamalini, Research Scholar,Computer Science Engineering, Bharath University,Chennai, India
Dr.M.Ponnavaikko, Provost, Vinayaka Mission's Research Foundation, AV Campus, Chennai, India.

competitors using competitive intelligence

d) discover the real time trending topics ie what people are talking about the industry, product, brand and customer opinions

e) keep track the virality of content spreads across the social media and world wide web.

Many Social Networks provides inbuilt Social Media Analytics tools on their Dashboards

a) Facebook Insights – gives complete statistics about your post, fans and reach on Facebook.

b) Instagram Insights – gives complete details about your profile, posts, stories, ads, and also gives detailed information about your followers such as their most active times and days.

c) Twitter analytics – used to analyse each Tweets and gives detailed information about retweets, likes, followers and followers' followers.

d) YouTube Analytics – used to understand the performance metrics like how people found your videos and how much they watched.

e) Google Analytics to see which social sites are referring the most traffic to your site. s primarily a web analytics tool, but it provides a small but important role in social media analysis: a breakdown of which social sites are driving traffic to your website.

Apart from inbuilt social media analytical tools various type of real time free tools like Followerwonk. ViralWoot, Tailwind, Keyhole and so on are used to analyse the different metrics of social media.

In order to find the various metrics of Social Media data, in this paper corpus of Twitter data is used. The results of these analysis explains how the social media data is important for business or a celebrity.

III. About Twitter

Twitter is the most popular microblogging site—one driven by short, textual messages or "microblogs." As of the end of 2011, 1 in 10 Internet users worldwide engaged with Twitter. Close behind LinkedIn, Twitter is the third most popular social network in the U.S. Twitter is often used as a place to report, react to, and engage with topics of national and international import, ranging from wars and natural disasters to elections and celebrity events. Twitter users can:

• Find and add friends. Adding friends is NOT a mutual relationship—they do not have to accept you as a friend for you to be able to follow them.

• Find and follow companies, entertainers, politicians, and more.

• Create a short bio—about one sentence in length. • Share links to anything on the Web.

• Use privacy settings to control what information is shared with whom.

• Track "trending topics"—the most popular topics of conversation on Twitter.

• Search for what all Twitter users are saying about a certain phrase, whether it is "trending" (very popular) or not

Tweet: A short, 140-character message Twitter users broadcast to their contacts. Twit/Tweeple/Tweeps: Nicknames for people who use Twitter. Retweet: A way to share another user's tweet with your own followers. @ Message: A way to mention or publicly message another Twitter individual. Mention: This is when someone @ messages your Twitter account. DM/Direct Message: A way to privately message another Twitter individual. Hashtags (#s): Denoted by a # in front of a word, hashtags are a way to link your tweet to an index of tweets on related topics. Ex: #NYC, #reading, #worldcup, #GOP, etc. Unfollow: This is when someone decides to remove a Twitter contact. The person will not be notified of this action. Favorite: If you like a tweet, then you can "favorite" it, and it will show up on your "Favorites" lists on your profile. The person whose tweets you like will also be notified. Lists/Listed: This is a way to organize the accounts you're following into categories. If you make your lists public, other people can follow them. Trends: This is a list of the top 10 phrases used on Twitter at any given moment. Microblogging: The act of broadcasting short, in-the-moment textual messages sent via platforms like Twitter

A. Volume of Tweets

Every second, on average [4], around 6,000 tweets are tweeted on Twitter (visualize them here), which corresponds to over 350,000 tweets sent per minute, **500 million tweets per day** and around 200 billion tweets per year.

From Twitter's launch in 2006 and until 2009, the volume of tweets grew at increasingly high rates, approaching a 1,400% gain in daily volume year to year [2]and around 1,000% gain in yearly volume. By mid 2010 the rate of growth started to cool down, sliding eventually below 100% gain in yearly volume in 2012. Today, the volume of tweets is growing at around 30% per year in our estimation.

B. Format of Tweet

In most of the social media, language used by the users is very informal. Users create their own words and spelling shortcuts and punctuation, misspellings, slang, new words, URLs, and genre specific terminology and abbreviations.

Twitter has developed its own language conventions [5]. The following are examples of Twitter conventions:

a) "RT" is an acronym for retweet, which indicates that the user is repeating or reposting.

b) "#" stands for hashtag is used to filter tweets according to topics or categories.

c) "@user1" represents that a message is a reply to a user whose user name is "user1".

d) Emoticons and colloquial expressions or slang languages are frequently used in tweets

e) External Web links (e.g. http://amze.ly/8K4n0t) are also frequently found in tweets to refer to some external sources.

f) Length: Tweets are limited to 140 characters.

The framework of Tweets Metrics Analyser is shown in figure below. It comprises of:

IV. Overview Of Tweets Metrics Analyzer

Tweets Metrics Analyzer is used to analysis the twitter data in an enhanced way and also explains how those data would have been used by business to enhance the customer experience. The framework of Tweets Metrics Analyser is shown in figure below. It comprises of:

- Tweets Extractor
- Enhanced Pre-processor
- Feature Extractor
- Metrics Analyser
- Metrics Visualizer.

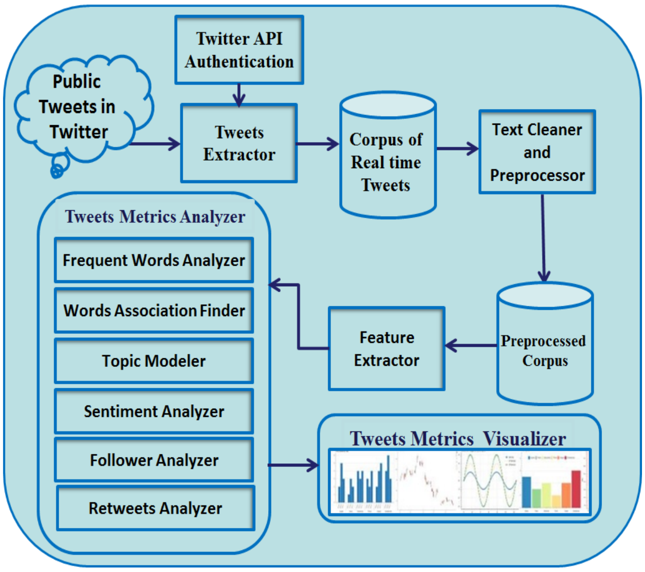


Fig.2. Tweets Metrics Analyzer with Enhanced Preprocessor

- Tweets Extractor: It is used to extract Tweets form Twitter after authenticating Twitter API.
- Enhanced Text Cleaner and Pre-processor: It is used to convert the raw text into clean text by removing numeric values, non-English characters, URLS, white spaces and stop words. It also handles case sensitive issues of text and stemming process.
- Feature Extractor: It is used to transform the tweets into set of features which represent the original data without any loss of information using dimension reduction technique.
- Tweets Metrics Analyser: It is used to analyse various aspects like the words which are frequently used in Tweets, Associations between other words, Topic Modelling, Sentiment Analysis, Flower Analysis and Retweet Analysis.
- Tweets Metrics Visualizer: It is used to visualize the output using visualization tools.

V. ENHANCED PREPROCESSOR AND TEXT CLEANER

A.Raw data

Tweets are slang words, which are used to express users' opinion or facts about current affairs in Twitter. People tweet personal messages, casual views, links, or anything that fits in 140 character requirements. The tweets consists of retweeted entry as RT, special characters, links, #tags, white spaces and stop words. These component will not add any value while analyzing the contents. These contents are cleaned and made fit for the analysis. The figure below shows the raw text extracted from twitter about GSAT II (Indian geostationary communications satellite).

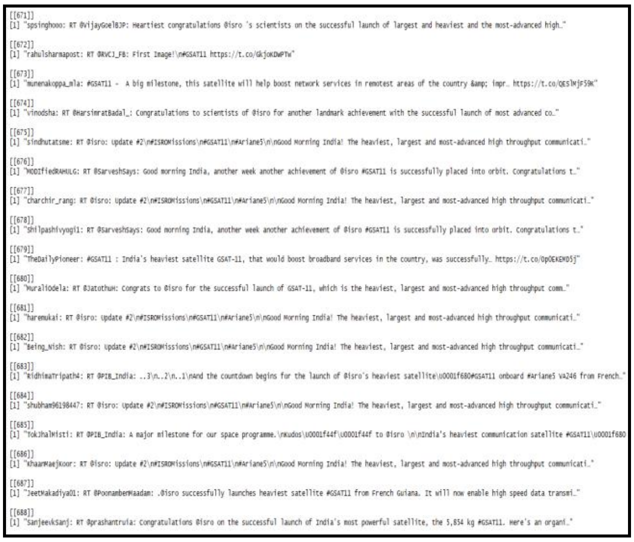

```
[[671]]
[1] "spsinghooo: RT @VijayGoelBJP: Heartiest congratulations @isro 's scientists on the successful launch of largest and heaviest and the most-advanced high…"

[[672]]
[1] "rahulsharmapost: RT @RVCJ_FB: First Image!\n#GSAT11 https://t.co/GkjoKDWPTw"

[[673]]
[1] "munenakoppa_mla: #GSAT11 -  A big milestone, this satellite will help boost network services in remotest areas of the country &amp; impr… https://t.co/QESlMjF59K"

[[674]]
[1] "vinodsha: RT @HarsimratBadal_: Congratulations to scientists of @isro for another landmark achievement with the successful launch of most advanced co…"

[[675]]
[1] "sindhutatsme: RT @isro: Update #2\n#ISROMissions\n#GSAT11\n#Ariane5\n\nGood Morning India! The heaviest, largest and most-advanced high throughput communicati…"

[[676]]
[1] "MODIfiedRAHULG: RT @SarveshSays: Good morning India, another week another achievement of @isro #GSAT11 is successfully placed into orbit. Congratulations t…"

[[677]]
[1] "charchir_rang: RT @isro: Update #2\n#ISROMissions\n#GSAT11\n#Ariane5\n\nGood Morning India! The heaviest, largest and most-advanced high throughput communicati…"

[[678]]
[1] "shilpashivyogi1: RT @SarveshSays: Good morning India, another week another achievement of @isro #GSAT11 is successfully placed into orbit. Congratulations t…"

[[679]]
[1] "TheDailyPioneer: #GSAT11 : India's heaviest satellite GSAT-11, that would boost broadband services in the country, was successfully… https://t.co/OpOEKEMD5j"

[[680]]
[1] "MuraliOdela: RT @JatothuH: Congrats to @isro for the successful launch of GSAT-11, which is the heaviest, largest and most-advanced high throughput comm…"

[[681]]
[1] "haremukai: RT @isro: Update #2\n#ISROMissions\n#GSAT11\n#Ariane5\n\nGood Morning India! The heaviest, largest and most-advanced high throughput communicati…"

[[682]]
[1] "Being_Nish: RT @isro: Update #2\n#ISROMissions\n#GSAT11\n#Ariane5\n\nGood Morning India! The heaviest, largest and most-advanced high throughput communicati…"

[[683]]
[1] "RidhimaTripath4: RT @PIB_India: ..3\n..2\n..1\nAnd the countdown begins for the launch of @isro's heaviest satellite\U0001f680#GSAT11 onboard #Ariane5 VA246 from French…"

[[684]]
[1] "shubham96198447: RT @isro: Update #2\n#ISROMissions\n#GSAT11\n#Ariane5\n\nGood Morning India! The heaviest, largest and most-advanced high throughput communicati…"

[[685]]
[1] "TokJhalMisti: RT @PIB_India: A major milestone for our space programme.\nKudos\U0001f44f\U0001f44f to @isro \n\nIndia's heaviest communication satellite #GSAT11\U0001f680

[[686]]
[1] "KhaarMaejKoor: RT @isro: Update #2\n#ISROMissions\n#GSAT11\n#Ariane5\n\nGood Morning India! The heaviest, largest and most-advanced high throughput communicati…"

[[687]]
[1] "JeetMakadiya01: RT @PoonamberMaadam: .@isro successfully launches heaviest satellite #GSAT11 from French Guiana. It will now enable high speed data transmi…"

[[688]]
[1] "SanjeevkSanj: RT @prashantruia: Congratulations @isro on the successful launch of India's most powerful satellite, the 5,854 kg #GSAT11. Here's an organi…"
```

Fig.3. Extracted Tweets about GSAT

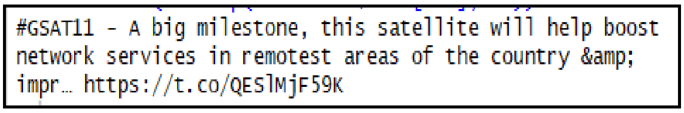

```
#GSAT11 - A big milestone, this satellite will help boost
network services in remotest areas of the country &amp;
impr… https://t.co/QESlMjF59K
```

Fig.4. Sample Tweet about GSAT11

B.Preprocessing

The data preprocessing can often have a significant impact on the performance of a supervised ML algorithm. The steps that are carried out by the enhanced preprocessor of this Tweet Metrics Analyzer are as follows:

a) Case Conversion: All words are converted either into lower case or upper case in order to remove the difference between "Text" and "text" for further processing.

b) Stop-words Removal: The commonly used words like a, an, the, has, have etc which carry no meaning i.e. do not help in determining the sentiment of text while analyzing should be removed from the input text.

c) Punctuation Removal: Punctuation marks such as comma or colon often carry no meaning for the textual analysis hence they can be removed from input text.

d) Stemming: Stemming usually refers to a simple process that chops off the ends of words to remove derivational affixes.

e) Lemmatization: Deals with removal of inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma.

f) Spelling Correction: Spelling of the incorrect words can be corrected based on automated selection of more probable word.

The steps which are used by the preprocessor to clean the raw is explained with the help of block diagram, which is given in figure below:

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

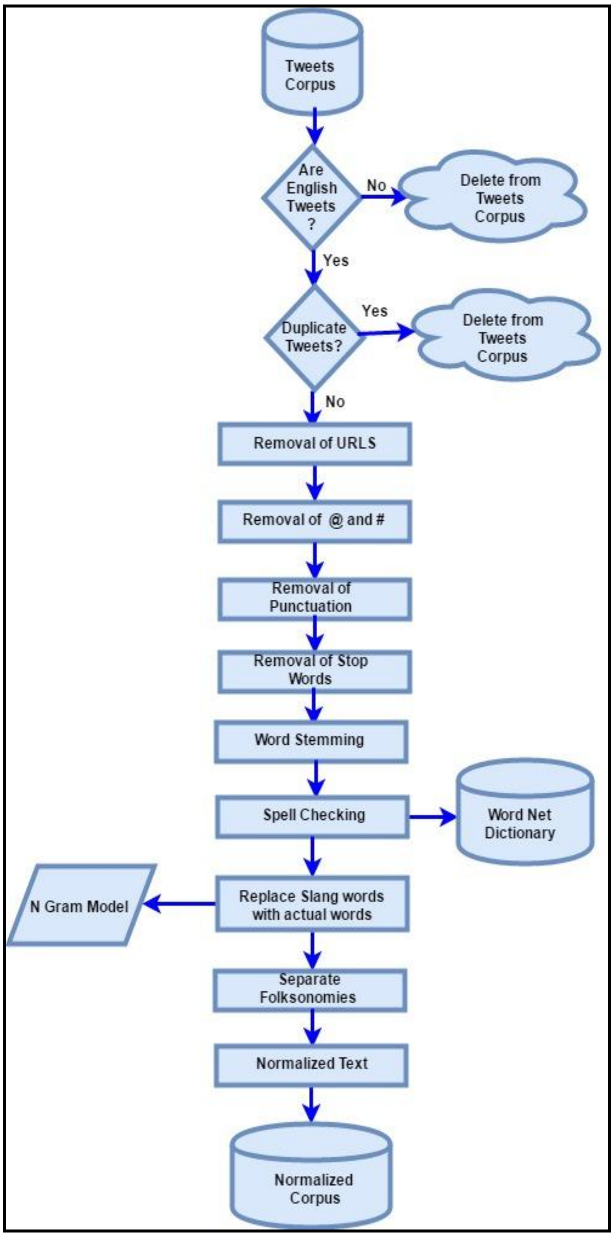


Fig.5. Enhanced text preprocessor

A.Remove Retweet Entities

Retweet feature Twitter i.e. re-posting of a Tweet helps the users to quickly share the Tweet with all of their followers. Users can retweet their own Tweets or Tweets from someone else. Sometimes users use “RT" at the starting of a Tweet to specify that they are re-posting somebody else's content. It increases the redundancy of data and will not add any additional value to the Data Source. The retweeted contents are removed for the data source in order to get the valid and unique contents. On applying retweet_Removal algorithm on the data source, out of 1500 tweets only 223 tweets were remaining which is depicted in figure 3 above.

B.Remove Html Links

A hyperlink is a word, phrase, or image that you can click on to jump to a new document or a new section within the current document. The user can type or paste the URL in the Tweet box on twitter. A URL of any length can be altered to length of 23 characters or less than 23 characters long. Hyper link contents also not valid and useful content for analysis of the content of the tweets. So removal of URL can add value to the analysis. The figure 4 above shows the contents of Tweet Database after removal of URL.

C.Remove Punctuations and Numbers

Numbers and punctuations have not added any addition meaning or value to the meaning of tweet text. That’s why they need to be removed from the content before using for analytics purpose.

D.Handle Cases(to lower case)

Text messages frequently have a variety of capitalization reflecting the starting of sentences and the proper nouns. The common approach is to convert the entire text to lower case for simplicity. While changing to lowercase, it is important to remember that words like “US” to “us”, can change meanings when changed to the lower case. The figure 6 above shows the contents of Tweet Database after changing the case.

E. Remove Stop words and white spaces

A words given in the text which are used to connect parts of a sentence rather than showing subjects, objects or intent. Word like “the” or “and” cab be removed by comparing text to a list of stop word.

VI. TWEETS METRICS ANALYSIS – METHODS AND ALGORITHMS

This section deal with methods, algorithms and techniques used to develop the tweets metrics analyser.

Twitter Authentication and Tweets Extraction

The R language packages “twitteR" and “ROAuth” are used to provide authentication to client application to access Twitter and extract tweets from the user timeline or based on search keyword. Sample tweet about “#GSAT” was extracted by this system for analytical purpose.

Text Cleaner and Preprocessor

Before performing any kind of analysis, the extracted tweets to be cleaned for accurate output. The “tm” package provides method for cleaning are removing unnecessary white space, punctuations, non-English characters, stop words and hyperlinks. Entire text to be converted into either uppercase or to lower case and applied to Stemming process to reduce the derived words to their root form such as making waking, walked into walk. The above processes have not add any value to the final results of the analysis.

Feature Extraction

Tweets contain the following fields [8]:

- ✓Created Date - UTC time (when the Tweet was created)
- ✓id– a sequence of integers to represent a unique key
- ✓id_str - string representation of the unique identifier
- ✓text - actual UTF-8 text posted by the user
- ✓source - utility used to post the Tweet,
- ✓truncated - indicates whether the value of the text parameter was truncated or not
- ✓user - user who posted this Tweet
- ✓coordinates - represents the geographic location of this Tweet

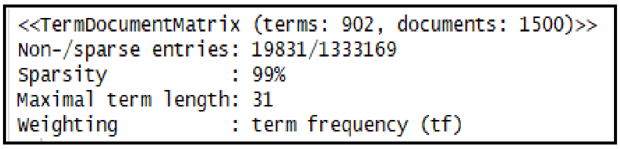
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

- ✓ place - indicates the place of Tweet
- ✓ retweet_count - the number of times Tweet has been retweeted

When the Tweets are extracted, all the above attributes are stored in the corpus. Not all the above attributes add value to the analysis. The feature extractor was used to identify the most relevant attributes that contain valuable information. Only "Text" field contains valid information and the feature extractor extracts that field for further processing. Method of dimension reduction is used by Feature Extractor.

Frequent Word Analyzer

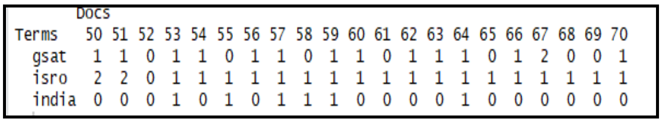
Frequent Word Analyzer[9] is used to find the most frequently used words in the corpus. The first step in the process of finding frequent words is the construction a Term Document Matrix (TDM). In the TDM, rows represent the documents, columns represent the terms and the matrix element value represent the number of occurrences of the term within a document. Frequent Word Analyzer considers each tweet as a document and counts the top occurring words stores them in term document matrix (TDM). Figure below showa TDM for 1500 documents and 902 terms is constructed i.e dimension of 1500 X 902 matrix is constructed with sparsity of 99%.



```
<<TermDocumentMatrix (terms: 902, documents: 1500)>>
Non-/sparse entries: 19831/1333169
Sparsity           : 99%
Maximal term length: 31
Weighting          : term frequency (tf)
```

Fig. 6. TDM – Term Document Matrix of GSAT Corpus

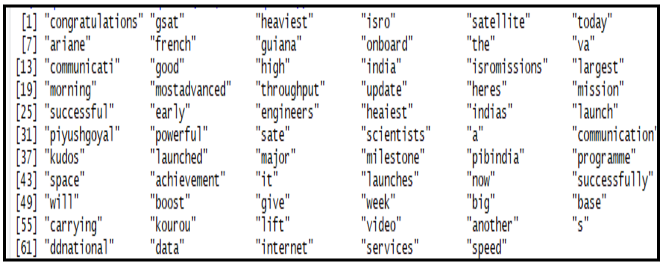
TDM is used to find the Top Frequent Terms of the Corpus. Figure below shows the occurrence of the terms gsat, isro and india in the documents from 50 to 70 in the corpus.



```
        Docs
Terms   50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70
  gsat   1  1  0  1  1  0  1  1  0  1  1  0  1  1  1  0  1  2  0  0  1
  isro   2  2  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  india  0  0  0  1  0  1  0  1  1  1  0  0  0  0  1  0  0  0  0  0  0
```

Fig. 7.Frequent occurrence of terms gsat, isro and india in documents 50-70 of Gsat Corpus

The figure below shows the top frequent words which occurs more than 500 times in gsat corpus.



```
 [1] "congratulations" "gsat"          "heaviest"    "isro"        "satellite"     "today"
 [7] "ariane"          "french"        "guiana"      "onboard"     "the"           "va"
[13] "communicati"     "good"          "high"        "india"       "isromissions"  "largest"
[19] "morning"         "mostadvanced"  "throughput"  "update"      "heres"         "mission"
[25] "successful"      "early"         "engineers"   "heaiest"     "indias"        "launch"
[31] "piyushgoyal"     "powerful"      "sate"        "scientists"  "a"             "communication"
[37] "kudos"           "launched"      "major"       "milestone"   "pibindia"      "programme"
[43] "space"           "achievement"   "it"          "launches"    "now"           "successfully"
[49] "will"            "boost"         "give"        "week"        "big"           "base"
[55] "carrying"        "kourou"        "lift"        "video"       "another"       "s"
[61] "ddnational"      "data"          "internet"    "services"    "speed"
```

Fig. 8. Top Frequent Terms of GSAT Corpus >=500 times

Graphical representation of Top Frequent words is shown below in figure.

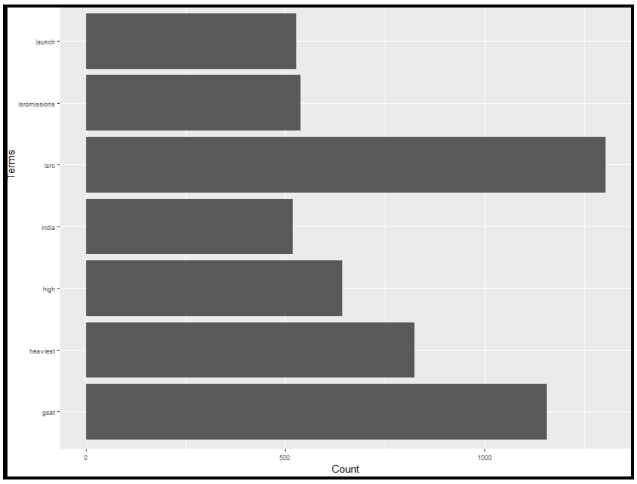


Fig 9. Graphical Representation of Top Frequent Words

Word cloud is a way to visually represent the frequent and important terms of a corpus. It is an image composed of words used in the documents of text corpus, wherein the size of each term indicates its significance and occurrence. Figure below shows the word cloud of top frequent words of the gsat corpus.

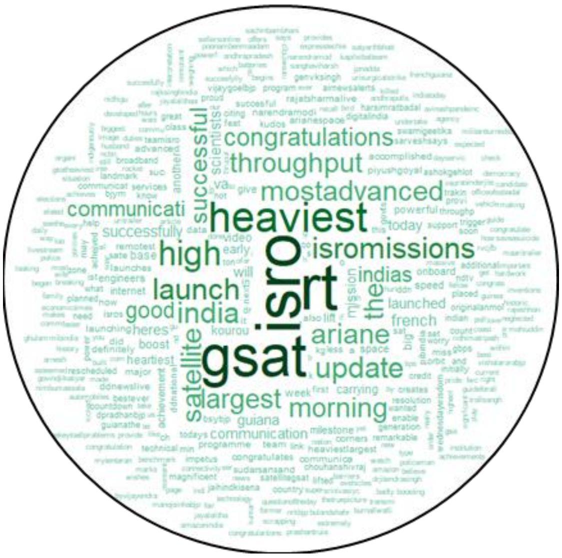
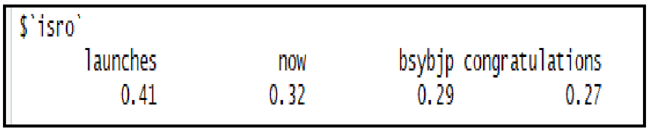


Fig 10 Wordcloud of Gsat Corpus

Word Association Finder

Word association Finder [10] is used to find the words which are associated with other words of the corpus. It also finds the percentage of association between them. The figures below shows the words which are associated with isro and gsat. The word "launches" has the highest association with the word "isro".

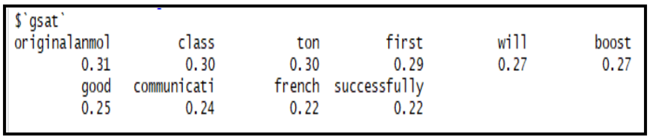


```
$`isro`
    launches          now     bsybjp congratulations
        0.41         0.32       0.29            0.27
```

Fig. 11. Words Associated with the word "isro"

The word "original anmol(Priceless)" has the highest association with the word "gsat".

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



```
$`gsat`
originalanmol         class           ton         first          will         boost
         0.31          0.30          0.30          0.29          0.27          0.27
         good   communicati        french  successfully
         0.25          0.24          0.22          0.22
```

Fig. 12. Words Associated with the word “gsat”

A. *Topic Modelling*

Topic modelling [11] is text-mining tool for finding the abstract "topics" that occur in a corpus of documents using probability theory. It is used to discover the hidden semantic structures in a text document. It simply used to clusters similar or associated words. Topic modeling is widely used in document classification, information retrieval, rating predictions and sentiment analysis. Latent Dirichlet allocation (LDA) is one of the most popular method for finding the hidden topic in the documents corpus. Topic Modeler uses LDA approach to find top 8 topics and first 5 terms of each topic. The figure below shows steam graph of top eight topics discovered in the gsat corpus using LDA approach.

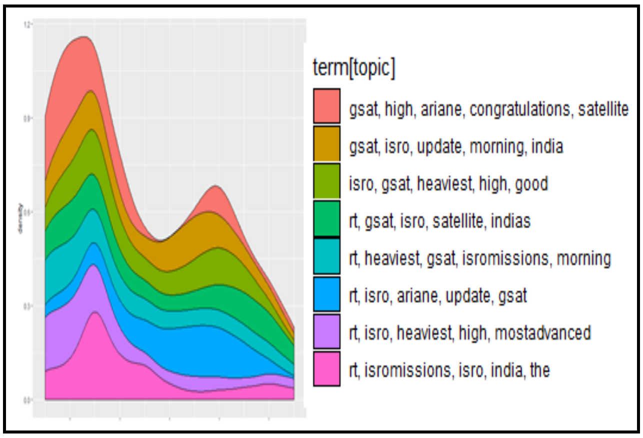


Fig. 13. Top eight Topics of Gsat Corpus

B. *Sentiment Analysis*

Sentiment Analyzer[11] is used to identify and categorize opinions and sentiments expressed by users in each tweet. Dictionary based approach was used to categorize the users’ opinions about a particular event, product, or person into positive, negative, or neutral. The figure below shows the classification of Gsat corpus into three categories positive, negative and neutral.

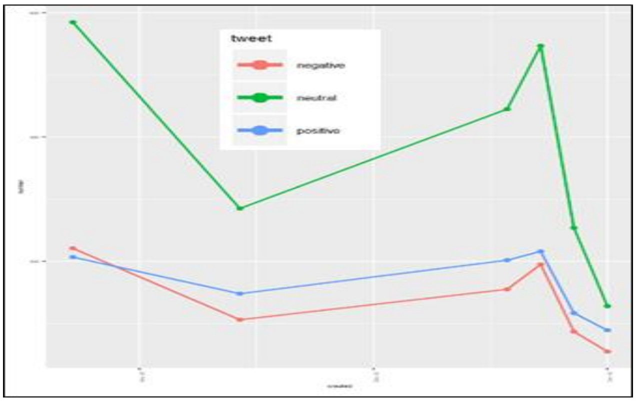


Fig. 14. Classifications of User Opinion about Gsat launch

But categorize tweets into positive, neutral or negative was not enough to find the exact polarity of a sentence. For example, the “good” has less rating than the word “better”. The affective lexicon based approach uses a special dictionary with rating of each word to find the rate of positivity and negativity in each tweet. Figure below shows the rate of positivity of each tweet in the Gsat corpus.

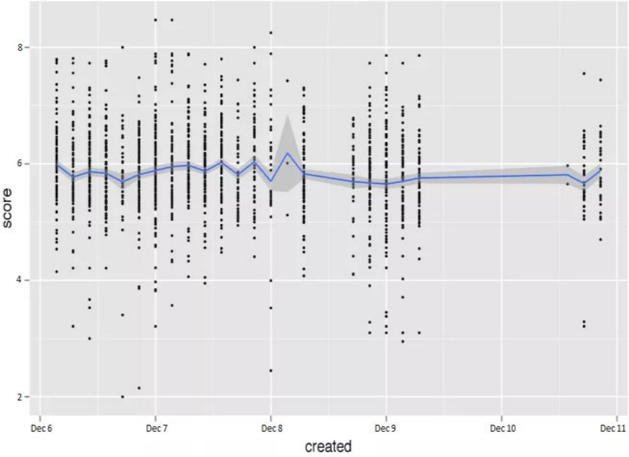
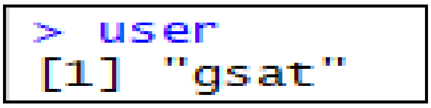


Fig. 15. Rate of Positivity of each tweet about Gsat launch

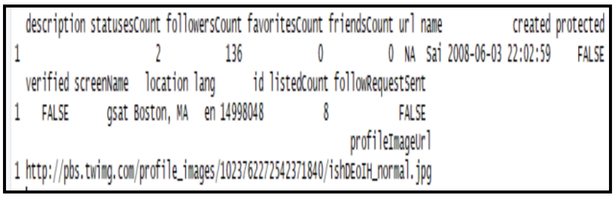
C. *User Info and Follower Analysis*

Follower Analysis is used to identify the influencers of the company, business or a brand. In ‘Twitter’, people can create the social relationships with other users by clicking ‘Follow’ option. Users can follow different posts of their interest. Same time different followers follow somebody in favor of them or just to support or criticize or to watch one’s behavior or simply tweets. The follower Analyzer used to find the various aspects of followers like active followers, their segments, best time to post the tweets and so on. Figure below shows the User Information and the statistics about the Followers.



```
> user
[1] "gsat"
```

Fig. 16. User info

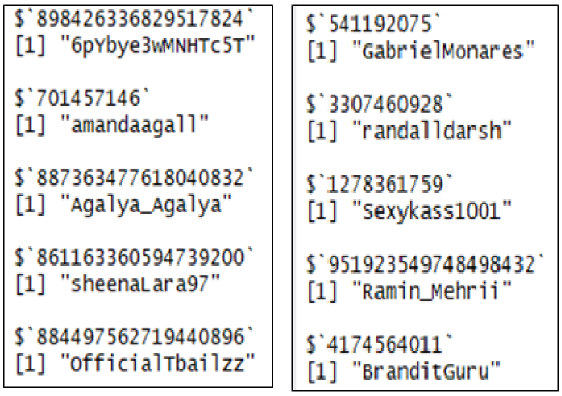


```
  description statusesCount followersCount favoritesCount friendsCount url name             created protected
1                         2            136              0            0  NA  Sai 2008-06-03 22:02:59     FALSE
  verified screenName   location lang        id listedCount followRequestSent
1    FALSE       gsat Boston, MA   en 14998048           8             FALSE
                                                                  profileImageUrl
1 http://pbs.twimg.com/profile_images/1023762272542371840/ishDEoIH_normal.jpg
```

Fig. 17. Follower Analysis – statistics

Follower’s details of user “gsat” is shown in figure below. There were 136 followers for the user “gsat”.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



```
$`898426336829517824`
[1] "6pYbye3wMNHTc5T"

$`701457146`
[1] "amandaagall"

$`887363477618040832`
[1] "Agalya_Agalya"

$`861163360594739200`
[1] "sheenaLara97"

$`884497562719440896`
[1] "officialTbailzz"
```

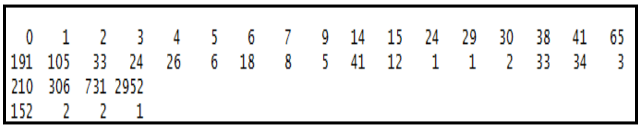
```
$`541192075`
[1] "GabrielMonares"

$`3307460928`
[1] "randalldarsh"

$`1278361759`
[1] "Sexykass1001"

$`951923549748498432`
[1] "Ramin_Mehrii"

$`4174564011`
[1] "BranditGuru"
```

Fig (a) Followers Fig (b) Follower's Follower

Fig. 18. Followers Analysis of User "Gsat"

D. *Retweet Analysis*

Retweet analysis of twitter gives important details about the customers to the business. If the content are retweeted by other users to their network means that users like you or your product or your brand. The figure below shows the tweet id and number of times that tweet is retweeted.



```
  0    1    2    3    4    5    6    7    9   14   15   24   29   30   38   41   65
191  105   33   24   26    6   18    8    5   41   12    1    1    2   33   34    3
210  306  731 2952
152    2    2    1
```

Fig. 19. Retweet Analysis

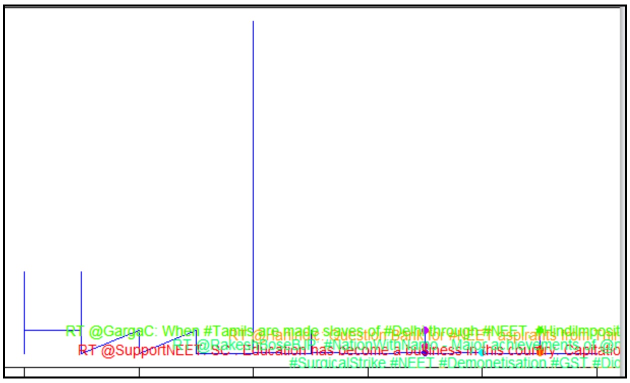


Fig. 19. Plot of Retweet Analysis

VII. CONCLUSIONS

This paper explains in detail about various aspects of Social media data. It also elaborates the need of the large volume of free social media data available online and finding different metrics like analysis of frequently used words, finding association between the words, topic modelling, sentiment analysis, follower analysis and retweet analysis. These metrics help the business to create successful social campaigns, recognize influencers for their brand, service & industry , compare strengths & weaknesses of competitors, discover the real time trending topics ie what people are talking about the business and customer opinions & sentiment towards their business.

ACKNOWLEDGMENTS

We wish to express our sincere thanks and deep sense of gratitude to all the staff members of computer department of Bharath University, Chennai for their support and co-ordination. We also wish to thank to our family members for their support to complete this work.

REFERENCES

1. https://en.wikipedia.org/wiki/Social_media accessed on Nov'2017.
2. https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/ accessed on Nov'2017
3. http://www.queenslibrary.org/sites/default/files/Social_Media/Twitter%20Tutorial.pdf accessed on Nov'2017
4. http://www.internetlivestats.com/twitter-statistics/ accessed on Nov'2017
5. Anuja P Jain, Asst. Prof Padma Dandannavar," Application of Machine Learning Techniques to Sentiment Analysis", Applied and Theoretical Computing and Communication Technology (iCATccT), IEEE 2016 ,pp. 628 – 632.
6. Raymond Kosala, Hendrik Blockeel," Web Mining Research: A Survey", ACM SIGKDD,Volume 2, Issue 1, July 2000.pp.1-15.
7. Shuyan Bai, Qingtian Han, Qiming Liu, Xiaoyan Gao "Research of an Algorithm Based on Web Usage Mining", Intelligent Systems and Applications International Workshop ,IEEE 2009, pp.1-4.
8. https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object(Accessed: Sep, 2018)
9. Dunđer, M. Horvat, S. Lugović, "Word occurrences and emotions in social media: Case study on a Twitter corpus", IEEE 2016, pp. 1284 – 1287.
10. Felipe Bravo-Marquez, Eibe Frank, Saif M. Mohammad and Bernhard Pfahringer," Determining Word–Emotion Associations from Tweets by Multi-Label Classification", 2016 IEEE/WIC/ACM International Conference on Web Intelligence,pp. 536 – 539.
11. Moh. Nasrul Aziz ; Ari Firmanto ; A. Miftah Fajrin ; R. V. Hari Ginardi," Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction", (ICITACEE), IEEE: 2018, pp. 125 – 130.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication