

Word Cloud for Online Mobile Phone Tweets towards Sentiment Analysis



Abstract: People became more eager to express and share their opinions on web regarding day-to-day activities or global issues. Social media contributed a transparent platform to share views across the world. Recently research communities, academia, public and service industries are working rigorously on sentiment analysis also termed as opinion mining, to extract and analyze public mood and views. Data pre-processing is a crucial step in sentiment analysis and selecting an appropriate pre-processing methods can improve classification accuracy. In this paper, we explore the role text pre-processing of online mobile phone reviews towards Sentiment Analysis. Proposed text pre-processing methods remove inconsistent and redundant elements on the collected data to improve classification accuracy. Proposed Pre-processing methods involves removal of punctuations, special characters, digits, escaping HTML characters, decoding data, Apostrophe Lookup, Removal of Stopwords, Removal of URLs, Removal of Expressions. The final pre-processed online review data are presented in the form of word cloud with the frequency statistics of the keywords.

Keywords: Text Pre-Processing, Sentiment analysis, Word Cloud, OnlineMobilePhone Reviews, Opinion mining, Accuracy.

I. INTRODUCTION

In the previous years, sentiment analysis has became a hottest topic in scientific and market oriented research in the field of Natural Language Processing and Machine Learning Techniques. Sentiment Analysis examines the problems of studying text like post and reviews, uploaded by users on microbloging platforms, forums and electronic business regarding the opinions they have about a product, services, events, person or idea etc. The basic uses of Sentiment Analysis to classify a text to positive or negative. Based on the different datasets, the opinion can be classified as binary that is positive or negative. Sentiment Analysis has gained popularity in Information Retrieval, Data mining, Text mining and computational linguistics in research organization for product reviews. Sentimental orientation is based on text classification which contains reviews and opinions. Sentiment analysis is computational study in which it contains opinions, sentiments, and emotions expressed in the text. For example, by obtaining consumer feedback on a marketing campaign, an organization can measure the campaigns success or learn how to adjust it for greater success. Feedbacks about product is also helpful in building good products compare to other competitors.

Revised Manuscript Received on October 30, 2019. * Correspondence Author

A.Kalaivani, Associate Professor, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science,(SIMATS).chennai

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an <u>open access</u> article under the CC BY-NC-ND license (<u>http://creativecommons.org/licenses/by-nc-nd/4.0/</u>)

There are mainly two methods to carry out the sentiment analysis, first is known as Supervised approach or Machine Learning based approach which make use of machine learning classification techniques and other is known as Unsupervised or Lexicon based approach, which is also known as dictionary based approach.Sentiment Analysis which is also known as Opinion Mining is a process of mining the user generated text ,content towards a product services from different social media. Opinions play a very vital role in decision making and are important for different organizations know that whether people like their products and services, what do people think about them, what kind of things people really like and dislike their product, service which may really help organizations to make decisions in a better way. But people are doing some product analysis before purchasing products. Some organization are conducting surveys and opinion polls from public which is expensive as well as time consuming. Sentiment analysis on twitter data and other social websites faces several challenges due to short messages and unstructured data. Data preprocessing methods play crucial step in sentiment analysis to preprocess the information which is necessary and make analysis to find out whether it is positive or negative. Various data preprocessing techniques will remove inconsistent and redundant data and visualize data based on most frequently used words using word cloud and word cloud2 techniques. The goal of present work to analyze different data pre-processing steps and to defines the best out of the considering methods. Therefore data pre-processing can be primary step in Sentiment Analysis , besides it is evaluated carefully , thus leaves an open questioning that why and to what extend does it increase the accuracy of the classifier.

The rest of the paper can be organized as follows. Section 2 provides some literature survey on data preprocessing. Section 3, on the other hand, presents block diagram. Section 4, methodology introduces our approach. Section 5 highlights the results at each and every phase of the preprocessing and visualization of the input tweets. Finally, section 6 presents the conclusions of the proposed work.

II. LITERATURE SURVEY

Different authors have done their research work in the field of Data pre-processing on different domains and proved good efficiency and accuracy in removing noisy data by different techniques. some of the authors mainly concentrated on stop word removal for achieving better accuracy. By using Ghag and Shah [1] observed data processing techniques on movie reviews for the effects of stop words removal .

Published By: Blue Eyes Intelligence Engineering & Sciences Publication



Retrieval Number F7964088619/2019©BEIESP DOI: 10.35940/ijeat.F7964.088619 Journal Website: www.ijeat.org

Naramula Venkatesh*, Assistant Professor, Vignana Bharathi Institute of Technology(VBIT),Ghatkesar

Accuracy on unprocessed dataset increased to removal stop words dataset by sentiment classifiers and display better than other classifier which is based on term weighting techniques.S. Rill et. al., in [2], they focused on system which is designed in detecting Political topic emerge in Twitter accounts. The basic idea is to develop relational graph for different emerging domain on political

information enriched like polarity and content. They have come up with Twitter hash tags special for sentiment hash tags in which people uses these tags to give opinions about leaders or leading parties for detection of polarity. They also improved feature based sentiment analysis method for knowledgeable base.F. H. Khan [3] et. al., Introduces a hybrid technique for sentiment determination and implemented for each tweet. The data preprocessing values are demonstrated using corrected and stop word removal method, detection and analyzed its abbreviations. They have used domain independent techniques to resolve data sparsely issues. Tested accuracy and compared with others to prove effectiveness of hybrid technique.Haddi et al in [4] applied a data pre-processing combination of and chi-square methods for removal of irrelevant features. The results showed an appropriate text pre-processing methods can significantly enhance the classifier's performance. The text pre-processing steps in sentiment analysis is to cleaning up the dataset from any noisy data, thus reducing the complexity of a document. Finally achieved good accuracy using SVM algorithm for clean datasets. Many online data contains uninformative parts and many noisy terms such as HTML tags which creates dimensionality problem and difficulty for the classification process. The methodologies which are used in cleaning the data and generating an informative data out of twitter messages may involve Punctuation and symbol removal, stop words removal, stemming as shown.Duwairi and El-orfali[5] conducted different pre-processing strategies on Arabic text as Dataset for defining opinions and reported positive results. The author mainly concentrated on removing of stop words which defines good accuracy on only Arabic text. Bhattacharajee[6] et. al., describe sentiment analysis using vector space model which is based on term frequency and conducted data preprocessing and filtered data to provide good data testing for rating .They also deals with TD-IDF value for defining whether terms are present in the document which is used finding co-efficient between different terms.

Poobana[7] et. al., gathered product reviews from various e-commerce websites and describe SVM algorithm for review rating. The system performed data processing on collected reviews in two stages for removing noise terms, first is stop words removal and other is through stemming for better results. Finally different authors used different techniques and algorithms in filtering noisy ,inconsistent data on different dataset and achieved appreciable results.

Basically the information received through online may contain variety of symbols, noisy information , uninformative sentences and parts like URLS, HTML tags etc. The words in the text may not have any impact and makes the dimensionality problem which leads difficulty for the classification of each word in the given text. Here preprocessing is required in order to remove all such noisy information , so than we can improve performance of the classification process which increase the speed of the classifier in real time sentiment analysis. Data preprocessing which involves in convert Raw information into readable format. Issues includes:

1.Incomplete

2.Inconsistant

3.Noisy

III. PROPOSED SYSTEM

In Real-world the data may contain unreadable format which lacking in certain trends ,unpolished ,disorder, and noisy with many errors. Data processing is the best model to resolve .A Proposed system shown in fig.1.works in three stages, first deals about Data collection from different social media and second , Data pre-processing in which the main challenges is to remove noisy data and uninformative format which includes HTML tags cleaning, punctuations and symbol removal stop words by using different methodologies out of twitter messages and finally visualized the most frequently used words for sentiment Analysis.



Fig.1. Proposed Data Pre-Processing System Block Diagram

3.1 Data Collection

Data collection is a process in which information can be collected on the specific topic or targeted variables. The data collected is a quantitative data or qualitative data. This can used to evaluate and answer some relevant questions in terms of a research or decision-making process. Many methods used to collect information in online social media for freely available information. The prior stage of data collection had a way in form of surveys, interviews and group discussions. Subsequently data can collected by using web Analytics tools. The collected data are in structured, semi-structured and unreadable manner. Some social media websites such as Facebook, Twitter, Blogs etc are accurate information from different APIs and places in table format as csv files. For the proposed work the dataset are collected from Twitter accounts.

3.2 Data Pre-Processing

Data Pre-Processing is a primary steps in the area of sentiment analysis and Opinion Mining.



Retrieval Number F7964088619/2019©BEIESP DOI: 10.35940/ijeat.F7964.088619 Journal Website: <u>www.ijeat.org</u>

684

Published By: Blue Eyes Intelligence Engineering & Sciences Publication



The Real world data doesn't make any sense as it is in unstructured, incomplete, noisy, and inconsistent and need to analyze using different type of data pre-processing techniques to discover knowledge data. Different types of Data preprocessing steps are:

1. Punctuation Removal: Stand alone punctuations, special characters and numerical tokens are removed as they do not contribute to sentiment which leaves only alphabetic characters.

This step needs the use of tokenized words as they have been split appropriately for us to remove.

2. Numbers Removal: This involves removing noise from text in its raw format which contain number in the data collected.

3. Removal of URL. Sometimes text data contains URL and hyperlinks for different reviews and comments which give in confusion of tokenization and information extraction

4. Removal of stopwords: Stop words are filtered out before further processing of text in which these words are contribute little to overall meaning and they are generally the most common words in a English language. For instance, "the," "and," and "a," while all required words in a particular passage, don't generally contribute greatly to one's understanding of content.

5. Removal of expression: This involves removing noise from text in its raw format. For example, the text is scrapped from the web it may contain HTML or XML wrappers or markups. Removal of these can be done through regular expressions. Fortunately our reviews do not apply to this as we were able to extract the exact review from the XML file.

3.3.Visualization

Visualization is an effective method for explore abstract ideas and also to communicate a knowledge information. For visualizing results for Sentiment Analysis, many different types of techniques are available such as graphs, histograms, and matrices. The most popular approaches used are Interactive Maps, Wordcloud etc.

VI. RESULTS AND DISCUSSION

4.1 Data Preprocessing

Different types of data pre-processing steps are implemented in order to remove all stopword, digits, punctuations marks, Alphanumeric characters from datasets. The primary technology used is R Tool. Data pre-processing can be implemented for the below five sample tweets which are noisy and inconsistent data and by filtering Irrelevant data such as hash tags,@,\$,!,stopwords by using stopwords removal, removing URLs, special characters, whitespaces 6. Removal of lowercase: The removal of lowercase involves avoids having multiple copies of same words. Next, we add some simple metrics for every text:

1.number of characters in the text 2.number of words in the text

• The next step consist in extracting vector representations for every review. The module Gensim creates a numerical vector representation of every word in the corpus by using the contexts in which they appear (Word2Vec). This is performed using shallow neural networks. What's interesting is that similar words will have similar representation vectors.

• Each text can also be transformed into numerical vectors using the word vectors (Doc2Vec). Same texts will also have similar representations and that is why we can use those vectors as training features.

• We first have to train a Doc2Vec model by feeding in our text data. By applying this model on our reviews, we can get those representation vectors.

Finally we add the TF-IDF (Term Frequency - Inverse Document Frequency) values for every word and every document.But why not simply counting how many times each word appears in every document? The problem with this method is that it doesn't take into account the relative importance of words in the texts. A word that appears in almost every text would not likely bring useful information for analysis. On the contrary, rare words may have a lot more of meanings.

Visualization methods are used in multimedia, medicine, education, engineering, science etc. The words with largest size is most frequently used and with less size are least used. By using such different size tell that customer is less or more discussed about product. So visualization will suggest analysts a better way to communicate valuable data in brief.

and expressions and keeping on data which are required for showing different form of data by using wordcloud and Barplot.

4.1.1 Removal of Punctuation marks

Punctuation are used for removing disambiguate on different sentences by spacing, conventional signs, and certain typographical. Figure 2 and Figure 3 shows the sample tweets before and after removal of punctuation marks.

Fig.1

1. RT @option_snipper: \$AAPL beat on both eps and revenues. SEES 4Q REV. \$49B-\$52B, EST. \$49.1B https://t.co/hfHXqj0IOB

685

5. \$AAPL - wow! This was supposed to be a throw-away quarter and AAPL beats by over 500 million in revenue! Trillion dollar company by 2018!



^{.2.} RT @option_snipper: \$AAPL beat on both eps and revenues. SEES 4Q REV. \$49B-\$52B, EST. \$49.1B https://t.co/hfHXqj0IOB 3. Let's see this break all timers. \$AAPL 156.89

^{.4.} RT @SylvaCap: Things might get ugly for apl with the iphone delay. With $apl down that means almost all of the FANG stocks were down posâ<math>d_1^i$

Word Cloud for Online Mobile Phone Tweets towards Sentiment Analysis

Fig.2 Tweets before removal of Punctuation Marks

- 1. rt optionsnipper aapl beat on both eps and revenues sees 4q rev 49b52b est 491b httpstcohfhxqj0iob
- 2. rt optionsnipper aapl beat on both eps and revenues sees 4q rev 49b52b est 491b httpstcohfhxqj0iob
- 3. lets see this break all timers appl 15689
- 4. rt sylvacap things might get ugly for aapl with the iphone delay with aapl down that means almost all of the fang stocks were down pos $\hat{a} \in \hat{C}$
- 5. aapl wow this was supposed to be a throwaway quarter and aapl beats by over 500 million in revenue trillion dollar company by 2018

Fig.3. Tweets after removal of Punctuation Marks

4.1.2. Removal of numbers

Twitter data contains some numbers which is not useful for analysis which can be removed for further analysis.

- 1. rt optionsnipper aapl beat on both eps and revenues sees 4q rev 49b52b est 491b httpstcohfhxqj0iob
- 2. rt optionsnipper aapl beat on both eps and revenues sees 4q rev 49b52b est 491b httpstcohfhxqj0iob
- [3] lets see this break all timers appl 15689

[4] rt sylvacap things might get ugly for aapl with the iphone delay with aapl down that means almost all of the fang stocks were down pos…

[5] aapl wow this was supposed to be a throwaway quarter and aapl beats by over 500 million in revenue trillion

Fig.4. Tweets before removal of numbers

1. rt optionsnipper aapl beat on both eps and revenues sees q rev bb est b httpstcohfhxqjiob

- 2. rt optionsnipper aapl beat on both eps and revenues sees q rev bb est b httpstcohfhxqjiob
- 3. lets see this break all timers aapl

4. rt sylvacap things might get ugly for aapl with the iphone delay with aapl down that means almost all of the fang stocks were down posâ€

5. aapl wow this was supposed to be a throwaway quarter and aapl beats by over million in revenue trillion dollar company by

Fig.5. Tweets after removal of numbers

4.1.3. Removal of URL

Sometimes text data contains url and hyperlinks for different reviews and comments .this can be removed from below .

[1] RT @option_snipper: \$AAPL beat on both eps and revenues. SEES 4Q REV. \$49B-\$52B, EST. \$49.1B https://t.co/hfHXqi0IOB [2] RT @option_snipper: \$AAPL beat on both eps and revenues. SEES 4Q REV. \$49B-\$52B, EST. \$49.1B https://t.co/hfHXqi0IOB [3] Let's see this break all timers. \$AAPL 156.89

[4] RT @SylvaCap: Things might get ugly for \$aapl with the iphone delay. With \$aapl down that means almost all of the FANG stocks were down pos…

[5] \$AAPL - wow! This was supposed to be a throw-away quarter and AAPL beats by over 500 million in revenue! Trillion dollar company by 2018!

Fig.6. Tweets before URL removal

:[1] rt optionsnipper aapl beat eps revenues sees q rev bb est b

[2] rt optionsnipper aapl beat eps revenues sees q rev bb est b

[3] lets see break timers aapl

[4] rt sylvacap things might get ugly aapl iphone delay aapl means almost fang stocks posâ ϵ_1^{l}

[5] aapl wow supposed throwaway quarter aapl beats million revenue trillion dollar company

Fig.7. Tweets after URL removal

4.1.4. Removal of stopwords

For data analysis at word level different words occurs repeatedly which are called stop word.

[1] rt optionsnipper aapl beat on both eps and revenues sees q rev bb est b httpstcohfhxqjiob

[2] rt optionsnipper aapl beat on both eps and revenues sees q rev bb est b httpstcohfhxqjiob

[3] lets see this break all timers aapl

[4] rt sylvacap things might get ugly for aapl with the iphone delay with aapl down that means almost all of the fang stocks were down posâ€

[5] appl wow this was supposed to be a throwaway quarter and appl beats by over million in revenue trillion dollar

Fig.8. Tweets Before removal of stop words



Published By:

& Sciences Publication



- [1] rt optionsnipper aapl beat eps revenues sees q rev bb est b httpstcohfhxqjiob
- [2] rt optionsnipper aapl beat eps revenues sees q rev bb est b httpstcohfhxqjiob
- [3] lets see break timers aapl
- [4] rt sylvacap things might get ugly aapl iphone delay aapl means almost fang stocks posâ ϵ_1^{l}
- [5] aapl wow supposed throwaway quarter aapl beats million revenue trillion dollar company

Fig.9. Tweets After removal of stopwords

4.1.5. Removal of expression and whitespaces :

Different word may contains some expression or certain pattern same.

- 1] rt optionsnipper aapl beat eps revenues sees q rev bb est b httpstcohfhxqjiob
- [2] rt optionsnipper aapl beat eps revenues sees g rev bb est b httpstcohfhxgjiob

[3] lets see break timers aapl

- [4] rt sylvacap things might get ugly aapl iphone delay aapl means almost fang stocks posâ€t
- [5] aapl wow supposed throwaway quarter aapl beats million revenue trillion dollar company

Fig.10. Tweets before removal of whitespaces:

- 1. rt optionsnipper beat eps revenues sees q rev bb est b
- 2. rt optionsnipper beat eps revenues sees q rev bb est b

3. lets see break timers

- 4. rt sylvacap things might get ugly iphone delay means almost fang stock posâ \in ¹
- 5. wow supposed throwawa

Fig.11. Tweets after removal of whitespaces

4.1.6. Converting uppercase into lowercase All uppercase letters are converted in lowercase letters.

Before converting lowercase:

1. RT @option_snipper: \$AAPL beat on both eps and revenues. SEES 4Q REV. \$49B-\$52B, EST. \$49.1B

https://t.co/hfHXqj0IOB

2. RT @option_snipper: \$AAPL beat on both eps and revenues. SEES 4Q REV. \$49B-\$52B, EST. \$49.1B https://t.co/hfHXqi0IOB 3. Let's see this break all timers. \$AAPL 156.89

4. RT @SylvaCap: Things might get ugly for \$aapl with the iphone delay. With \$aapl down that means almost all of the FANG stocks were down posâ€

5. \$AAPL - wow! This was supposed to be a throw-away quarter and AAPL beats by over 500 million in revenue! Trillion dollar company by 2018

Fig.12. Tweets before converting uppercase into lowercase

1. rt @option_snipper: \$aapl beat on both eps and revenues. sees 4q rev. \$49b-\$52b, est. \$49.1b https://t.co/hfhxqj0iob

- 2. rt @option_snipper: \$aapl beat on both eps and revenues. sees 4q rev. \$49b-\$52b, est. \$49.1b https://t.co/hfhxqj0iob 3. let's see this break all timers. \$aapl 156.89

4. rt @sylvacap: things might get ugly for \$aapl with the iphone delay. with \$aapl down that means almost all of the fang stocks were down posâ€

5. \$aapl - wow! this was supposed to be a throw-away quarter and aapl beats by over 500 million in revenue! trillion dollar company by 2018!

7.calculating Term document matrix: During pre-processing process we define occurrence of different words in the form of matrix called as Term Document Matrix:

Fig.13. Tweet after converting uppercase into lowercase

4.1.7.Term-frequency matrix

In this matrix it display all terms present in twitter data with their frequency indicating no.of times each word is used in the documents.



Word Cloud for Online Mobile Phone Tweets towards Sentiment Analysis

Do	cs																				
Terms	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
beat	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
eps	1	1	0	0	0	1	1	1	0	2	0	0	0	1	0	0	1	1	0	0	
est	1	1	0	0	0	2	2	0	0	0	0	0	0	2	0	0	2	2	0	0	
optionsnipper	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
rev	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	
revenues	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
sees	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
break	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
lets	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
see	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Fig.13. Tweet after converting uppercase into lowercase

4.1.8. Visualization of wordcloud

The below are frequently used word from twitter datasets in which all the words(>25), which is input for wordcloud and Barplot.

eps	est	almost	delay	fang	get
94	136	88	86	87	135
iphone	means	might	posÃ	stock	sylvacap
139	86	95	85	136	170
things	ugly	cnbcnow	earnings	revs	breaking
89	86	29	383	55	34
expected	amp	reports	sales	tonight	cnbc
32	42	110	32	33	28
big	move	call	like	market	qqq
42	25	29	45	25	33
ready	stocktwits	amzn	waiting	calls	rather
36	48	28	28	108	86
heres	today	dont	just	live	will
31	107	92	36	57	54
report	billion	now	yahoofinance	month	year
31	25	28	32	37	40
time	anything	close	earth	expect	headÃ
26	88	108	85	88	85
moving	puts				
88	93				

Fig.15. Word Cloud Formation

4.2 Visualization

The system uses different datasets from twitter and many text-Preprocessing techniques are applied and visualize the data using wordcloud and wordcloud2 which are basic packages for visualizing datasets on sentiment analysis. Each word importance can be interpretive graphical by using word cloud with different sizes as a cloud. The words with largest size is most frequently used and with less size



Fig.16. Bar Plot for Frequent Words

least used. By using such different size tell that are customer is less or more discussed about product. Barplot can be plotted by using x-axis which represents frequently used words in datasets and y-axis represents no. of times each word is used.



Fig.17. Word Cloud for Frequent Words



Retrieval Number F7964088619/2019©BEIESP DOI: 10.35940/ijeat.F7964.088619 Journal Website: www.ijeat.org

688

Published By:

& Sciences Publication





Fig 18.1.wordcloud2 formation as star

6. Conclusion

Sentiment analysis has become both challenging field with lot of difficulties in natural language processing. Different application will get advantage through its results in the form of news analytics, questioning and answering, marketing soon. The main challenge of

ssentiment analysis is to build human understandable system. In this paper we have identified sentiment of sproduct reviews. Different pre-processing method are used to reduce noise text and also visualize data using wordcloud. In addition, visualizing the classification results in the form of rating chart based on opinions about product by customers to capture and assessment about any products. Future scope this pre-processed document can be input for any machine Learning techniques to classify opinions.

ACKNOWLEDGEMENT:

The author would thank to Higher Authorities of SIMATS(Saveetha University) for use of Research Laboratories. I greatefully acknowledge the computing resources provided by DIST-FIST lab in VBIT to carried out the work for publishing the paper.

REFERENCES:

- K. V. Ghag and K. Shah, "Comparative analysis of effect of stop words removal on sentiment classification," in IEEE International Conference on Computer Communication and Control (IC4-2015), (2015), pp. 2–7.
- S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, "PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis," Knowledge-Based Systems, 69(1), pp. 24-33, (2014).
- 3. F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining
- Frame work using hybrid classification scheme", Decision Support Systems, 57(1), pp. 245-257, (2014).
- 5. E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-processing in Sentiment Analysis," in Procedia Computer Science, , vol. 17, pp. 26–32 (2013).
- R. Duwairi and M. El-orfali, "A Study of the Effects of Preprocessing Strategies on Sentiment Analysis for Arabic Text," J. Inf. Sci., pp. 1– 13, (2013).
- Bhattacharjee, S., Das, A., Bhattacharya, U., Parui, S. K., & Roy, S. 2015, Sentiment analysis using cosine similarity measure, In Recent Trends in Information Systems (ReTIS), IEEE 2nd International Conference on pp. 27-32(2015).
- SashiRekha k., Poobana S , Opinion Mining from Text Reviews Using Machine Learning Algorithm, International Journal of Innovative Research in Computer and Communication Engineering Vol. 3 Issue 3 pp.1567-1570.(2015).

Retrieval Number F7964088619/2019©BEIESP DOI: 10.35940/ijeat.F7964.088619 Journal Website: <u>www.ijeat.org</u>



Fig 18.2.wordcloud2 with frequently words

AUTHORS PROFILE:



N.Venkatesh currently working as Asst.Professor in VBIT,Ghatkesar,pursing Phd in Saveetha University(SIMATS).Chennai.

Dr.A.Kalaivani completed her Phd from Anna University in Image Processing.She was award title "South Indian Best Academic Faculty Award From SIAA 2018.She is currently working as Associate.Professor in SIMATS,Chennai,TM.



Published By: Blue Eyes Intelligence Engineering & Sciences Publication