January-February 2020
ISSN: 0193-4120Page No. 6756 - 6763

Sentiment Emotions Scoring for Apple Mobile Tweets

[1]R. Vijayalakshmi, [2]A. Kalaivani

[1]Final Year CSE Student, [2]Associate Professor,
[1,2]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai
[1]vijayalakshmir449@gmail.com, [2]kalaivanischolar@gmail.com

Article Info
Volume 82
Page Number: 6756 - 6763
Publication Issue:
January-February 2020

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 01 February 2020*

Abstract

Sentimental analysis is one of the hottest topic on social media and used in web applications. Sentiment analysis is all about expressing their emotions and opinions on a particular topic or on the particular product in the form of reviews and tweets. Sentiment analysis can express their emotions in the form of positive, negative or in the form of neutral. The product performance or product usage can be known through sentimental analysis. In our proposed system, the dataset chosen for sentiment analysis is apple mobile phone dataset one is before earnings and the other one is after the earnings. The datasets is applied to pre-processing strategies to remove inconsistent and redundant factors. The proposed methods of pre-processing include the deletion of punctuations, special characters, numbers, HTML characters escaping. The dataset is further fine tuned by applying decoding data, Apostrophe Lookup, removing stop words, removing URLs and eliminating expressions. Visualization of the preprocessed information is represented as word cloud with the frequency facts of the key words. Finally the tweets are classified into emotions based on nrc_sentiment dictinory and descriptive analysis for the emotions.

Keywords: Sentiment analysis, Mobile tweets, nrc_sentiment dictionary, Sentiment Score.

1. Introduction

In the preceding years, sentiment analysis has have become a hottest subject in medical and market oriented studies in the field of Natural Language Processing and Machine Learning Techniques. Sentiment Analysis explores the troubles of reading text like tweets and reviews, uploaded by customers on micro blogging platforms, forums and electronic business. The opinions can be approximately on product, services, events, individual or concept etc. The fundamental uses of sentiment Analysis is to categorise a textual content in the form of emotions like positive or negative. Sentiment Analysis has earned reputation in Data mining, Information Retrieval, Text mining in studies enterprise for product opinions. This process is based on text type which contains evaluation and critiques. Sentiment analysis is computational to take a look at where it contains evaluations, sentiments, and emotions expressed within the text. For instance, consumer comments on advertising a product an company can degree the product success or learn the way to modify it for greater fulfilment. Feedbacks are also used to find the approximately product is also useful when constructing excellent merchandise relative to different rivals. In particular, there can be two techniques for performing sentiment evaluation. First is known as the Supervised Technique or Machine Learning-based Method which makes use of gadget learning category strategies and specific is known as the Unsupervised or Lexicon-based Method, also known as the primarily based dictionary method. Sentiment Analysis that's also known as Opinion Mining is a technique of mining the consumer created text information from special social media toward a product services. Opinions play a very important role in making choices and are crucial for different organizations to identify that or not human beings like their merchandise and offerings, what do human beings consider them, what kind of factors human beings obviously like and hate their product, a carrier that can genuinely help companies to create higher choices. But humans do some product analysis before they buy products. Some organizations engage in public surveys and opinion polls that are expensive as well as time consuming. Due to quick texts and unstructured data, sentiment assessment on twitter facts and other social web sites faces various challenges. Data pre-processing strategies play an important role in the analysis of sentiments in pre-processing the data that is essential and analyzing whether it is high quality or negative. Different pre-processing statistics methods are inconsistent and unnecessary facts, and use word cloud and phrase cloud2 methods to visualize statistics based on the most commonly used words. The intention of present work to investigate special data pre- processing steps and to defines the satisfactory out of considering techniques. Therefore, pre-processing information may be number one phase in sentiment analysis, besides being carefully assessed, leaving an open wondering that is why and to what extent it improves the classifier's accuracy. The remainder of the paper can be written as follows. In section 2 affords literature survey on data pre- processing. Section 3 explains the block diagram of the proposed system and section 4 describes the methodology of the proposed system. Finally section 5 highlights the outcomes at each and every section of the pre- processing and visualization of the input tweets. Finally, section 6 gives the conclusion of the proposed work in the form of graphs that specifies the emotion levels of the tweets.

2. Literature Survey

Various authors have executed their research work inside the subject of data pre-processing on one kind of domain names and proved suitable performance and precision in getting rid of noisy information with the aid of distinctive strategies. Many of the authors specially focused on stop phrase removal for achieving higher precision. Naramula Venkatesh and A. Kalaivani [1], proposed a word cloud formation on apple mobile phone data sets after applying pre-processing and finding the frequency of the words. Bhattacharajee [2] et.al., describes sentiment evaluation using vector space model which is based on term frequency. They performed pre-processing statistics and filtered information to give proper rating data. They also discuss with TD-IDF charge for outlining what terms are present in the file used to find co-green between different phrases. Ghag and shah [3] observed statistics processing strategies on movie evaluations for the results of stop phrases elimination. Using sentiment classifiers and showing higher than other classifier based entirely on term weighting strategies, precision on unprocessed data sets extended to elimination for stall words dataset. S. Rill et. Al., in [4], they targeted on device that is designed in detecting Political subject matter emerge in Twitter accounts. The basic concept is to build relational graph on political news enriched like polarity and information for a specific emerging domain. They've given you special Twitter hash tags for sentiment hash tags where people use these tags to offer leadership evaluations or leading polarity detection events. They also enhanced function for knowledge base primarily based on sentiment evaluation method. F. H. khan [5] et. Al., Introduces a hybrid sentiment technique that will power and be implemented for each tweet.

Pre-processing values for tweets are shown using the right and stop word elimination system, detecting and analyzing its abbreviations. They used domain-independent methods to sparsely solve the issue of data. Tested precision and compared with others to show hybrid method's effectiveness. Haddi et al implemented a mix of pre-processing facts and chi-square techniques to remove irrelevant aspects in [6]. The results indicate pre-processing steps in the evaluation of sentiment is to clean up the dataset from any loud information, thus lowering a report's difficulty and achieving right precision using SVM algorithm for clean data set. Many online data incorporates uninformative components and several noisy terms like HTML tags that generate dimensionality disorder for the process in the category. Methodologies that can be used to clean statistics and generate informative statistics from twitter messages may also involve punctuations and removing symbols, removing words, and stemming words. Duwairi and El-orfali [7] conducted unique pre-processing technique pre-processing techniques on Arabic textual content as dataset for defining opinions and reported high quality results. The author has specially focused on preventing terms that describe correct accuracy on best textual content in Arabic. The device executed two ranges of data processing on collected critics. Firstly, it prevents the reduction of phrases and secondly, it removes the stemming of phrases for better results. Finally extraordinary authors used various techniques and algorithms in filtering noise, contradictory facts on special datasets and done appreciable results. Basically the data acquired via online can also contain style of symbols, noisy statistics, uninformative sentences and elements like ULRS, HTML tags etc. The terms inside the text will have no effect and will cause the problem of dimensionality within the given text for the form of each sentence. Here pre-processing is needed on the way to put off all such noisy information, so than we are able to increase the efficiency of the category manner which increases the velocity of the classifier in actual time sentiment analysis. Pre-processing data that includes translating the raw facts into readable format.

3. Proposed System

Data’s can be gathered by using web analytics tools which are independent, semi-based and unreadable manner. Social media web sites include Facebook, Twitter, Blogs and so forth are precise data from extraordinary APIs and location in desk format as csv files. Data Pre- Processing is a number one steps within the area of sentiment analysis and opinion mining.

The facts of the real world are useless as they are in unstructured, incomplete, noisy and inconsistent, and to be implemented and inconsistent, different types of data processing strategies are applied to discover information documents. Various types of data pre-processing steps are Punctuation, Number, URL, stop words, expression, lowercase removal. Removal process includes punctuations, special characters, URL and hyperlinks and numerical tokens. Stop words removal includes words such "the," "and" and "a" and removal of expression eliminates noise from text in its raw format. Lower-case elimination avoids having more than one copy of the same words. Visualization is one of the effective method for discover abstract thoughts and express to a expertise data. There are many distinct types of techniques available for visualizing outcomes for Sentiment Analysis, including graphs histograms and matrices. The most famous words used are Interactive Maps, Word cloud etc. Visualization strategies are in multimedia, medicine, education, engineering, technological know how etc. The words with the largest size are most frequently used and the least used with much less length. By the usage of such distinctive length inform that consumers is less or extra discussed approximately product. Thus, visualization will support analysts in a better way to briefly communicate valuable records.

Figure 1: Proposed System for Sentiment Scores on Mobile Phone Tweets

4. Results and Discussion

On the way to putting off all stop word, digits, punctuation marks, alphanumeric characters from datasets, various types of data pre-processing steps are applied. Data pre-processing may be carried out for the given sample five tweets which might be noisy and inconsistent. The tweets are taken off with the aid of filtering irrelevant records inclusive of hash tags,@,$,!,stop words by way of the use of stop words removal, removing URLs, unique characters, whitespaces and expressions. The pre processed data are depicted in various form of data by using word cloud and bar plot.

Removal of Punctuation Marks

Punctuation is used by spacing, traditional signs and sure typography to avoid disambiguate on exclusive sentences. Figure 2 shows the Apple Mobile phone Tweets sample and Figure 3 shows the punctuation marks after elimination.

January-February 2020
ISSN: 0193-4120Page No. 6756 - 6763

[1] RT @philstockworld: Whipsaw Wednesday - The View from Dow 22,000#Dow22000 $DIA $DXD $AAPL #Oil #Dollar -- https://t.co/e0RxvmkO6D https://…
[2] RT @philstockworld: Whipsaw Wednesday - The View from Dow 22,000#Dow22000 $DIA $DXD $AAPL #Oil #Dollar -- https://t.co/e0RxvmkO6D https://…
[3] RT @stockpicklist: $NBDR is making moves toward #triple #digit #gains Make your move #today #StockMarket #paid $psid $aapl $fb $aghi $ottv…
[4] RT @philstockworld: Whipsaw Wednesday - The View from Dow 22,000#Dow22000 $DIA $DXD $AAPL #Oil #Dollar -- https://t.co/e0RxvmkO6D https://…
[5] RT @TDAJJKinahan: With $AAPL behind us, focus turns to TSLA, with hopes of more insight into the company's delivery schedule.https://t.co/…

Figure 2: Sample Apple Mobile Phone Tweets

[1] rtphilstockworld whipsaw wednesday the view from dow 22000 dow22000 diadxdaapl oil dollar httpstcoe0rxvmko6dhttps…
[2] rtphilstockworld whipsaw wednesday the view from dow 22000 dow22000 diadxdaapl oil dollar httpstcoe0rxvmko6dhttps…
[3] rtstockpicklistnbdr is making moves toward triple digit gains make your move to day stockmarket paid psidaaplfbaghiottv…
[4] rtphilstockworld whipsaw wednesday the view from dow 22000 dow22000 diadxdaapl oil dollar httpstcoe0rxvmko6dhttps…
[5] rttdajjkinahan with aapl behind us focus turns to tsla with hopes of more insight into company's delivery schedulehttpstco…

Figure 3: Tweets after change case and removal of punctuation marks

Removal of numbers and URL

This data includes certain numbers that are not useful for analyzing and can be omitted for further study. In some cases the text data includes url and hyperlinks for various reviews, and comments can be omitted from below.

[1] rtphilstockworld whipsaw wednesday view dowdowdiadxdaapl oildollar …
[2] rtphilstockworld whipsaw wednesday view dowdowdiadxdaapl oildollar …
[3] rtstockpicklistnbdr making moves toward triple digit gains make move today stockmarket paid psidaaplfbaghiottv…
[4] rtphilstockworld whipsaw wednesday view dowdowdiadxdaapl oildollar …
[5] rttdajjkinahanaapl behind us focus turns tsla hopes insight company's delivery schedule …

Figure 4: Tweets after number and URL removal

[1] rtphilstockworld whipsaw wednesday view dowdowdiadxdaapl oil dollar…
[2] rtphilstockworld whipsaw wednesday view dowdowdiadxdaapl oil dollar…
[3] rtstockpicklistnbdr making moves toward triple digit gains make move today stockmarket paid psidaaplfbaghiottv…
[4] rtphilstockworld whipsaw wednesday view dowdowdiadxdaapl oil dollar…
[5] rttdajjkinahanaapl behind us focus turns tsla hopes insight company's delivery schedule…

Figure 5: Tweets after removal of whitespaces

Term-Frequency Matrix

The term frequency matrix shows all terms in twitter information and their frequency represents the number of times each word is used in the documents. The term frequency in the datasets represents x-axis, and y-axis represents number of times each word is used.

Terms	1 23	45	6 7	8	9	10	11	12	13	14	15	16	17	18	19	20
aapl	1 1 1	1 1	1 1	1	1	1	1	1	1	1	0	1	1	1	1	1
dia	1 1 0	1 0	0 0	0	1	0	0	1	1	0	0	1	0	1	0	0
dollar	1 1 0	10	0 0	0	1	0	0	1	0	0	0	1	1	1	0	0
dow	2 2 0	2 0	0 0	0	2	0	0	2	0	0	0	2	0	2	0	0
dxd	1 1 0	1 0	0 0	0	1	0	0	1	0	0	0	1	0	1	0	0

Figure 6: Term Frequency matrix

Visualization

The system made use of unique datasets from twitter and several text pre-processing methods are carried out and visualizes the statistics the usage of word cloud and word cloud2 which can be fundamental packages for visualizing datasets on sentiment analysis. Using phrase cloud with exclusive sizes as a cloud, each phrase importance can be interpretive graphical. The phrases with biggest length is most regularly used and with less length are least used. Use x-axis, which represents commonly used terms in datasets, bar plot can be plotted and y-axis represents no. times a word is used. Based on the below bar graph word cloud for the words present in the tweets are shown in figure8.

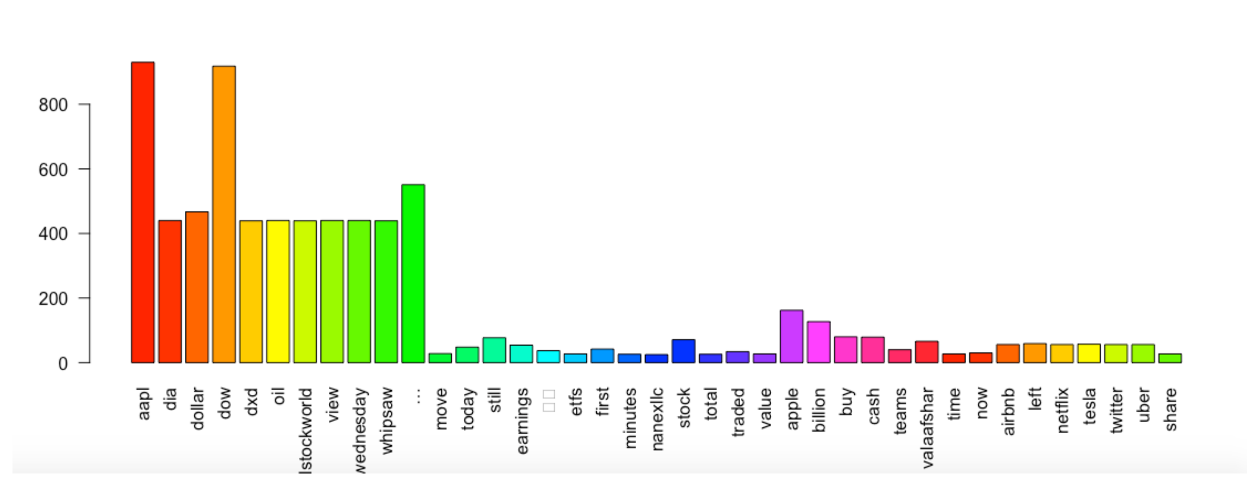

Figure 7: Bar Plot for Frequent Word

Figure 8: Word Cloud for Frequent Words

This sentiment scores are obtained from an nrc_sentiment dictionary that dictionary contains set of word for some emotions like anger expectation, disgust, fear, joy, sorrow, surprise, faith, negative, positive by using this we can able to provide a score for sentiment.

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	1	0	0	0	0	0	0	0	2

Figure 9: Sentiment Scores

Finally we can get the bar graph for this above mentioned emotions by that apple phone datasets before earnings shown in figure 10 and the other bar graph is after earnings shown in figure 11.

TEST Engineering & Management

January-February 2020
ISSN: 0193-4120Page No. 6756 - 6763

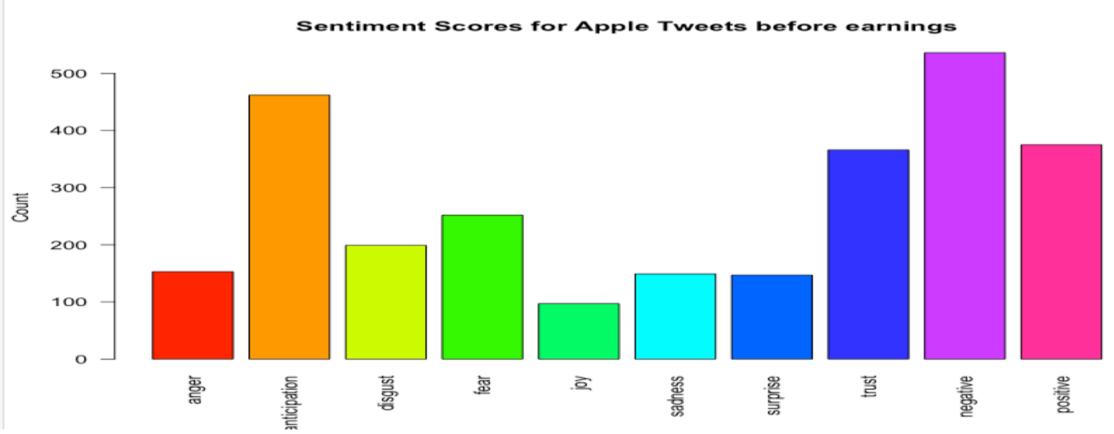

Figure 10: Sentiment Scores for Apple Tweets before Earnings

If we see the above bar graph before earnings the tweets are negative about the apple phone. People might think the product quality is not good and the cost may be more.

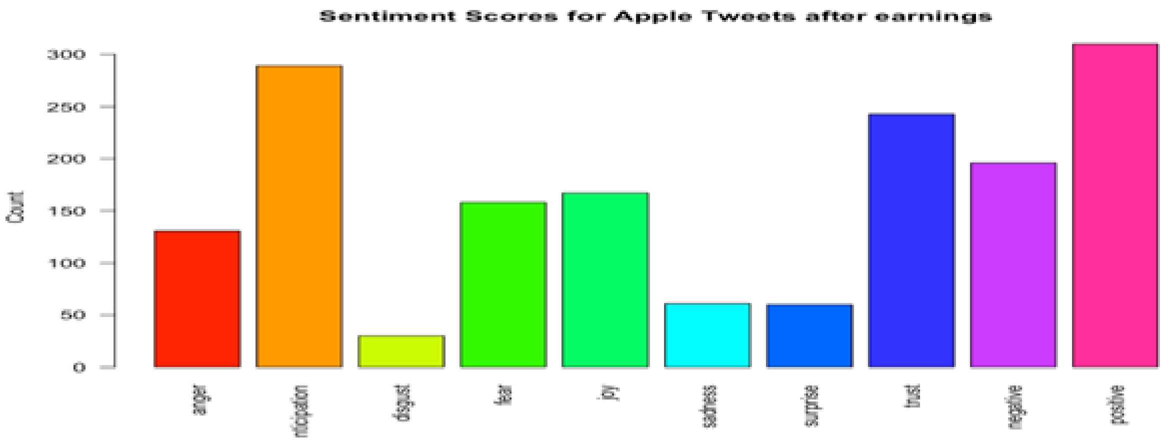

Figure 11: Sentiment Scores for Apple Tweets after Earnings

While coming to the tweets after earnings of apple phone the tweets are positive because the quality of the product is good and it is reasonable in price.

5. Conclusion

Evaluation of sentiment has grown to be both difficult area with many problems in language processing and the main venture of evaluation of sentiment is to construct a comprehensible human system. We have identified product review feeling in this paper. Different methods of pre-processing are used to reduce textual noise content and also to visualize documents using word cloud. Furthermore, the classification effects are visualized in the form of a score chart based entirely on reviews of approximately product customers to capture and analyze approximately all products Future scope this pre-processed file can be input for any system, learning techniques to classify critiques.

References

[1] A. Kalaivani, N. Venkatesh, "Word Cloud for Online Mobile Phone Tweets towards Sentiment Analysis" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019

[2] 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.

[3] K. V. Ghag and K. Shah, "Comparing analysis of effect of stop words removal on sentiment classification," in IEEE International Conference on Computer Communication and Control (IC4-2015), pp. 2–7, 2015.

[4] S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, "PoliTwi: Early detection of emerging Politics topic on twitter and the impact on concept-level sentiment analysis," Knowledge-Based Systems,

69, pp. 24-33, 2014.
[5] 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.
[6] 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.
[7] 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
[8] Hussain A., Razak H.A., Mkpojiogu E.O.C. (2017). The perceived usability of automated testing tools for mobile applications. Journal of Engineering Science and Technology. Vol 12, Special Issue 4. Page 89-97