

A Framework for Measuring Emotional Responses of Individuals in Social media by using Big data Approach

Dr. P. Sathishkumar¹, Dr. R Suguna², S Deepa³ and Dr M S Arunkumar⁴

¹⁻² Assistant Professor, Department of CSE, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamilnadu
Email: sathishkumarp@bitsathy.ac.in, sugunar@bitsathy.ac.in

³ Assistant Professor, Computer Technology, Kongu Engineering College, Perundurai, Tamilnadu, India
sdeepakec@gmail.com

⁴ Associate Professors, Department of CSE, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai
arunkumarmsster@gmail.com

Abstract—Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. It is possible to analyze the data and get answers immediately with today's technology. Sentimental Analysis is used to analyze opinions of the individual or group related to a business, product or service. Sentimental analytics is often driven by an algorithm along with voice inflection to identify the feelings of people regarding the topic. Also, the amount stress in the customer's voice and how fast they are speaking. In social media, the difficulties faced by the analyst is language most of the people doesn't use the proper format (like how r u, gud mrg etc.). To overcome such situation sentimental analytics device uses the Natural Language Processor (NLP). Another difficulty is if a word that actually carries the positive meaning but due to sarcastic manner could easily be misinterpreted. The paper mainly focuses on analysis of the text of the individual's emotions like joy, surprise, anger, disgust, positive and negative in social media and offers the result.

Index Terms— Big Data Analytics- Sentimental Analysis-NLP- Customer-Algorithm.

I. INTRODUCTION

Big data analytics helps organizations harness their data and use it to identify new opportunities and also leads to smarter business moves, more efficient operations, higher profits and happier customers. The most important values of Big Data are Cost reduction, Faster, better decisionmaking and new products and services. Big Data Analytics largely involves collecting data from different sources, munge it in a way that it becomes available to be consumed by analysts and finally deliver data products useful to the organization business. The process of converting large amounts of unstructured raw data, retrieved from different sources to a data product useful for organizations forms the core of Big Data Analytics. The main scope of this work is to analyse the emotions of the people in social media. Identifying how they are reacting to the situations and calculating the mentality of the people. It is done by taking the text from the social media. Converting the text document into data set and then by applying the algorithm to the data set we will get the required result consisting of each emotion with their

related percentages. From this result, we can analyse how they are reacting to the topic and what is their opinion regarding the topic. Then, this project provides the statistics for the emotions.

A. Structure of the Paper

The paper is organized in a following way. Chapter 1 covers the introduction part and chapter 2 briefs about the related works, chapter 3 defines the proposed methodology, chapter 4 describes the design and results analysis and chapter 5 provides the conclusion and future work of the proposed work.

II. LITERATURE REVIEW

Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. Starting from being a document level classification task (Turney, 2002; Pang and Lee, 2004), it has been handled at the sentence level (Hu and Liu, 2004; Kim and Hovy, 2004) and more recently at the phrase level (Wilson et al., 2005; Agarwal et al., 2009). Microblog data like Twitter, on which users post real time reactions to and opinions about “everything”, poses newer and different challenges. Some of the early and recent results on sentiment analysis of Twitter data are by Go et al. (2009), (Bermingham and Smeaton, 2010) and Pak and Paroubek (2010). Go et al. (2009) use distant learning to acquire sentiment data. They use tweets ending in positive emoticons like “:)” “:-)” as positive and negative emoticons like “:(” “:-(” as negative. They build models using Naive Bayes, MaxEnt and Support Vector Machines (SVM), and they report SVM outperforms other classifiers. In terms of feature space, they try a Unigram, Bigram model in conjunction with parts-of-speech (POS) features. They note that the unigram model outperforms all other models. Specifically, bigrams and POS features do not help. Pak and Paroubek (2010) collect data following a similar distant learning paradigm. They perform a different classification task though: subjective versus objective. For subjective data they collect the tweets ending with emoticons in the same manner as Go et al. (2009).

For objective data they crawl twitter accounts of popular newspapers like “New York Times”, “Washington Posts” etc. They report that POS and bigrams both help (contrary to results presented by Go et al. (2009)). Both these approaches, however, are primarily based on ngram models. Moreover, the data they use for training and testing is collected by search queries and is therefore biased. In contrast, we present features that achieve a significant gain over a unigram baseline. In addition we explore a different method of data representation and report significant improvement over the unigram models. Another contribution of this paper is that we report results on manually annotated data that does not suffer from any known biases. Our data is a random sample of streaming tweets unlike data collected by using specific queries.

The size of our hand-labeled data allows us to perform cross validation experiments and check for the variance in performance of the classifier across folds. Another significant effort for sentiment classification on Twitter data is by Barbosa and Feng (2010).

III. ANALYSIS AND DESIGN

We have taken text from the whatsapp group. Export the chat from group and send to the mail. Then save the text in notepad or wordpad in .txt file extension.

- Set the R studio directory using `setwd()` function.
- Then load the package “syuzhet”. Which contains package to load the document file and to run the sentiment function.
- Variable “text” is used to store the text document.
- `Get_nrc_sentiment()` function is used for sentiment analyse.
- `View()` is used to view the analyzed data in tabular form.
- `Barplot()` function is used to represent the outcome in bar chart.
- `Pie()` function is used to represent the outcome in pie chart.

IV. RESOURCES AND PREPROCESSING OF DATA

In this paper we introduce two new resources for pre-processing WhatsApp data: 1) an emoticon dictionary and 2) an acronym dictionary. We prepare the emoticon dictionary by labeling 200 emoticons listed on Wikipedia¹ with their emotional state. For example, “:)” is labeled as positive whereas “:(” is labeled as negative.

We assign each emoticon a label from the following set of labels: Extremely-positive, Extremely-negative, Positive, Negative, and Neutral. We compile an acronym dictionary from an online resource.² The dictionary has translations for 6,000 acronyms. For example, lol is translated to laughing out loud

V. THE PROPOSED METHODOLOGY

The Proposed Methodology involves number of steps. First we need to create a WhatsApp group followed by exporting of chats happened in the whats App group. Then the data will be stored in the database and algorithm is applied to extract and store the analyzed data and the results will be displayed.

The parameters to be considered for analyzing the tweeted message angry, surprise, happiness, sad emotions. The related words with perspective to these parameters will be identified and stored in database. Then the statements in whatsapp group will be fetched and given as an input and based on these stored related words ,the system recognizes and provides the count against with parameters.

The process related to the proposed methodology is shown in pictorial representation in Figure.1.

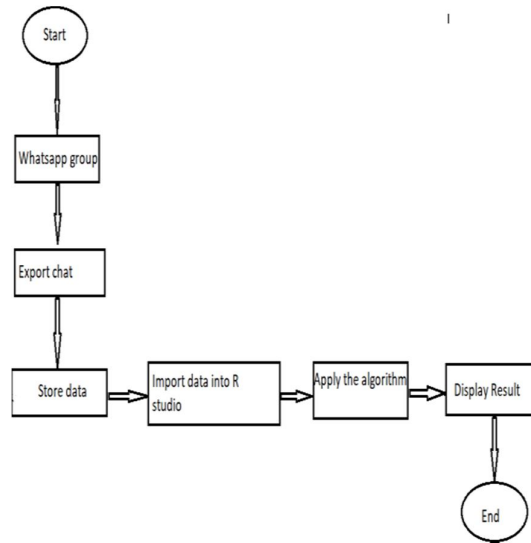


Fig.1 The proposed methodology

VI. RESULTS AND DISCUSSION

This chapter clearly shows the simulation results obtained for the proposed work. Fig.2 and 3 depicts the tabular view of the tweeted results

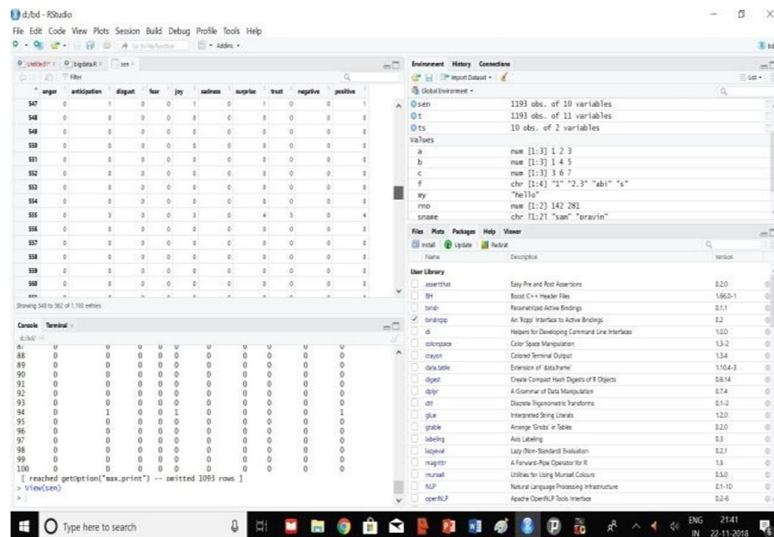


Fig.2 Table view of output-2

The fig.2 shows the tweeted messages related to various emotions such as surprise, angry, happiness and sad are shown in a form of pie chart.

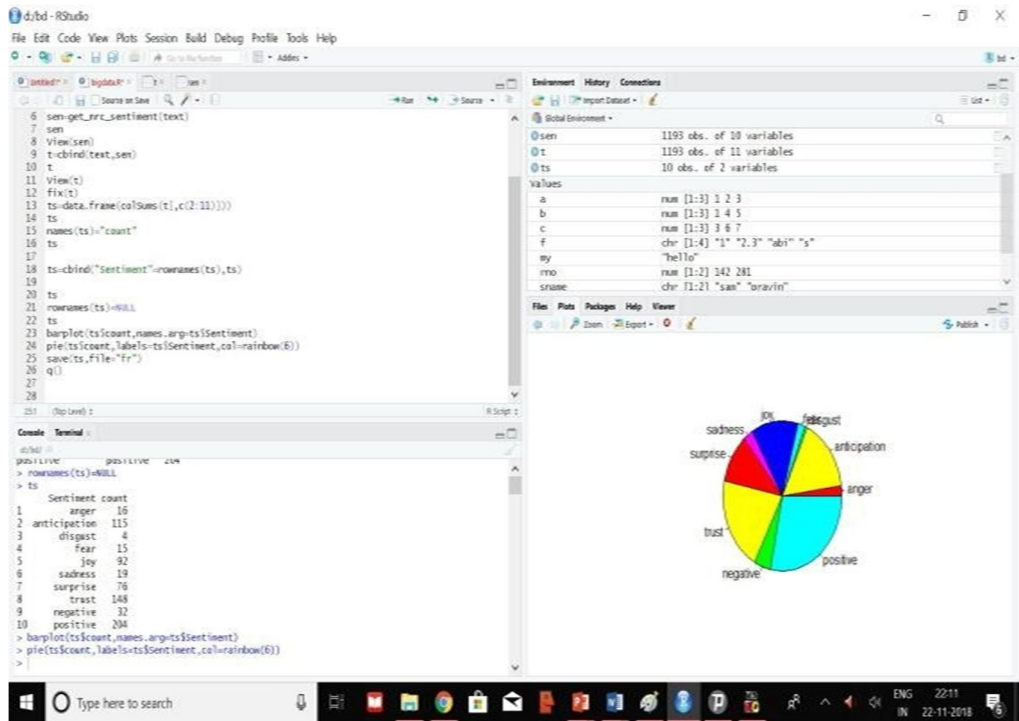


Figure 2: Result shown in pie chart

The fig.3 shows the tweeted messages related to various emotions such as surprise, angry, happiness and sad are shown in a form of bar chart.

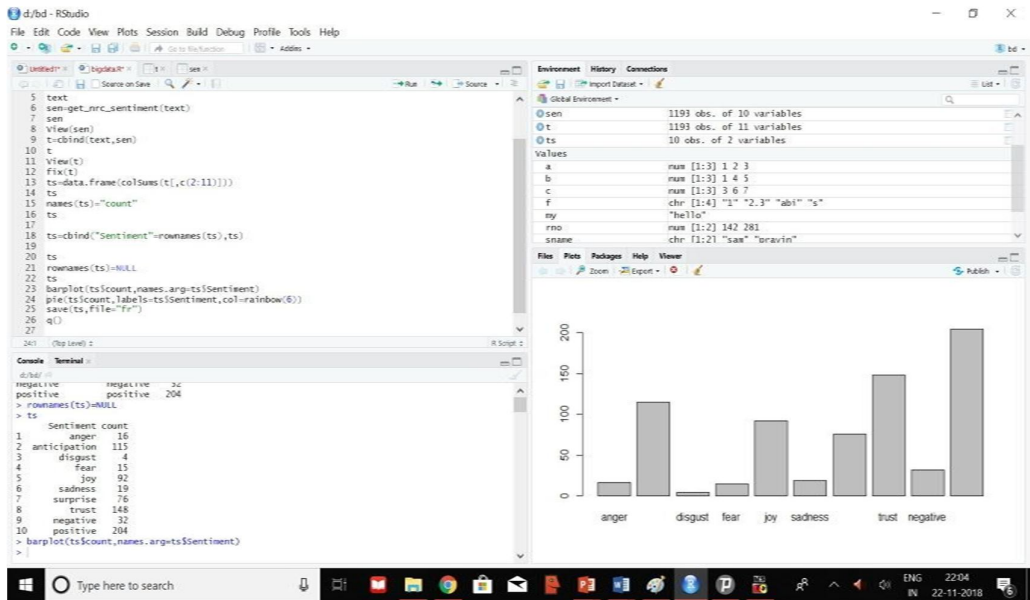


Figure 3: Result shown in Gantt chart

The fig.4 shows the individual counts of the tweeted messages against the various emotions such as surprise, angry, happiness and sad are shown in a form of bar chart.

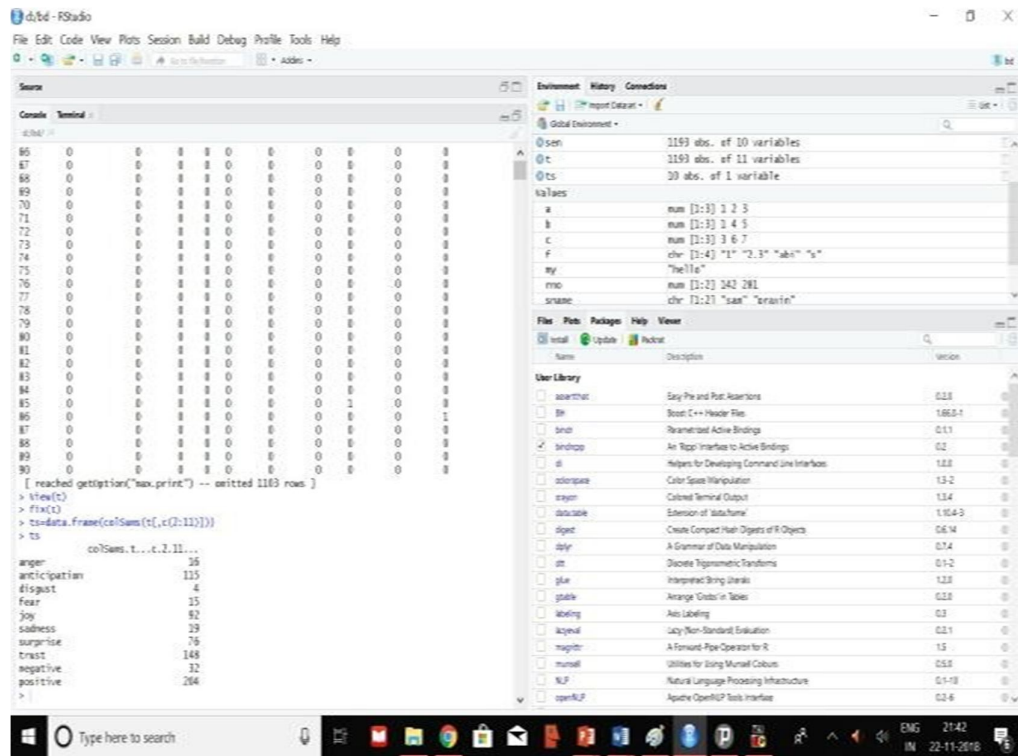


Figure 4: Result shown in Count

VII. CONCLUSION AND FUTURE WORKS

In this paper, words in the uploaded social media document are categorized based on their emotions and their counts are viewed in the console page. In other console page, the emotions are plotted in the bar graph, pie graph and line graph. Based on the graph and result, the behaviour and emotions of the social media users' can be analysed. This analysis will also be helpful for the release of many new products and brands in the real world. Sentiment analysis is only available for the English words so it does not able to understand other language words. So, in future, the sentiment analysis will be updated to accept more languages in addition to English and this algorithm can also be used for more social media. The future of sentiment analysis goes to still dig deeper, way past the surface of the amount of likes, comments and shares, and aim to achieve, and really perceive, the importance of social media interactions and they tell shoppers behind the screens.

REFERENCES

- [1] Apoorv Agarwal, Fadi Biadisy, and Kathleen Mckeown. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 24–32, March.
- [2] Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44.
- [3] Kim S-M, Hovy E (2004) Determining the sentiment of opinions In: Proceedings of the 20th international conference on Computational Linguistics, page 1367.. Association for Computational Linguistics, Stroudsburg, PA, USA.
- [4] Adam Birmingham and Alan Smeaton. 2010. Classifying sentiment in microblogs: is brevity an advantage? ACM, pages 1833–1836.
- [5] C. Fellbaum. 1998. Wordnet, an electronic lexical database. MIT Press. Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. Proceedings of the 20th international conference on Computational Linguistics.
- [6] Liu B, Hu M, Cheng J (2005) Opinion observer: Analyzing and comparing opinions on the web In: Proceedings of the

- 14th International Conference on World Wide Web, WWW '05, 342–351.. ACM, New York, NY, USA.
- [7] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford. David Haussler. 1999. Convolution kernels on discrete structures. Technical report, University of California at Santa Cruz.
 - [8] M Hu and B Liu. 2004. Mining and summarizing customer reviews. KDD. S M Kim and E Hovy. 2004. Determining the sentiment of opinions.
 - [9] Coling. Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pages 423–430.
 - [10] Turney PD (2002) Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, 417–424.. Association for Computational Linguistics, Stroudsburg, PA, USA.
 - [11] Alessandro Moschitti. 2006. Efficient convolution kernels for dependency and constituent syntactic trees. In Proceedings of the 17th European Conference on Machine Learning.
 - [12] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of LREC.
 - [13] B. Pang and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity analysis using subjectivity summarization based on minimum cuts.
 - [14] ACL. P. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. ACL.
 - [15] C M Whissel. 1989. The dictionary of Affect in Language. Emotion: theory research and experience, Acad press London. T. Wilson, J. Wiebe, and P. Hoffman. 2005. Recognizing contextual polarity in phrase level sentiment analysis. ACL.