

A Survey on Sentiment Analysis with Twitter

Palak Arora¹, Shivani Garg²

¹Student, Shri Ram College of Engineering & Management, Palwal, Haryana, 121102, India

²Assistant Professor, Aravali College of Engineering & Management, Greater Faridabad, Haryana, 121006, India

Abstract

As we know, the social networking sites such as Facebook, Twitter are becoming more popular and their usage is increasing rapidly everywhere, which gives everyone a fair chance to give their views on how they really feel about the different things out there. The perspective given by different users can be used as a tool to bring out their sentiment and thoughts regarding the items or public services and they might use this data for better business and the market in upcoming time. This feedback from users is very essential and also makes the process of decision making very much easier, but it includes some complications in terms of data management in order to identify wrong or right user's opinions.

This paper is based on the sentiment analysis design, i.e. pulling out a large number of tweets. Sentiment analysis we can also say, as open mining is something which helps to know the point of view behind beliefs written on various topics. Here we are going to use an open source methodology to visualize user's beliefs with R Programming (Statistical tool).

We have used 8 different feelings i.e. anger, disgust, anticipation, fear, sadness, surprise, trust with two special sentiments, which expresses people's good and bad thoughts.[1]

Keywords: Twitter; Sentimental; NLP; RStudio; Word Cloud; Politics awareness;

1. Introduction

With the upcoming time,[2] the pace of technology is at boom. Development had resulted in the adjustment of individual methods to communicate their views and beliefs. We are aware of the existence of approximately 111 small web logging websites. Small web logging sites are only internet social network WebPages where clients make brief, regular posts. Twitter is a popular micro-blogging service that allows you to read and post comments that make 148 char long. Twitter messaging is likewise named as Tweets. Tweets can be utilized as natural information. We will utilize a strategy that naturally picks up tweets in neural, negative and positive feelings. With the use of sentimental analysis, the client may find out the review about item or services prior to making a purchase. The corporation could use sentiment analysis to understand how customers feel about their products, allowing them to analyze client experience therefore; as a result, they can improve the product. Currently, approximately 6500 tweets get sent every second, resulting in approximately 561.6 million twitter posts a day. For the most part, these streams of tweets are chaotic, representing multi-point, fluctuating frame of mind in an unprocessed and unorganized fashion. Visualizing unorganized information is a difficult process in and of itself, and pulling out valuable data from that too is even more difficult. We need technologies and advanced tools that can help with tweet management and pulling out data from them. In spite of the fact that there are different routes possible to do it, as such, the activity is carried out in the paper using R Programming language.

Sentiment Analysis is the strategy for determining if a piece of written content is good, bad or impartial. This importantly involves exploring the emotions regards to penning on any subject. Sentiment model intends to analyze people's feelings, beliefs, perspectives, or preferences. We commonly classify the emotion in numerous categories in this model. Introduce the paper, and put a nomenclature if necessary, in a box with the same font size as the rest of the paper. The paragraphs continue from here and are only separated by headings, subheadings, images and formulae. Thesection headings are arranged by numbers, bold and 10 pt. Here follows further instructions for authors.[3]

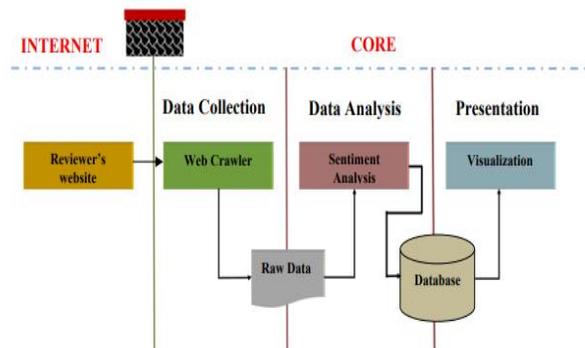


Fig1. Sentiment Analysis Model

1.1. Four Layers to which this analysis can be implemented:

There are four layers to which this analysis could be implemented: [4]

- Layer 1 is indeed the phrase level, where each sentence recognizes bad, neutral, or positive attitudes.
- Layer 2 acknowledges the full record of sentiments like a single unit or item, whether good, bad, or impartial.
- Layer 3 is the aspect layer, which is used when traits are present within an item, document, or entry phrase. If taken into account, the aspect level can lead to better analysis and findings. On the aspect level, in which all are applied, several sentiment analysis strategies have been implemented to clustering. All qualities with a sentiment result that is similar are clustered altogether.
- Layer 4 is the customer level, which applies graph theory to manage the societal relationships between different customers.

1.2. Need of Sentiment analysis :

Due to the growth of internet life, sentiment analysis has become especially critical. Sentiment analysis could be used to a wide range of activities and approaches. Several of those have been involved in the following fields:

Business: - Companies in the advertising industry use it to figure out ways, comprehend consumers' perception of a product or brands, what people respond to ads or product releases, or why people aren't buying certain things. Data Collection INTERNET CORE Data Analysis Presentation Reviewer's website Sentiment Visualization Analysis Web Crawler Raw Data Database 3.Politics: - It is used in the political world to analyze political ideas, as well as to identify contradictions including inconsistencies between administration acts and pronouncements. This could also be used to forecast election results! Public Acts: Analysis is often used to track and measure sociological phenomena, identify potentially risky circumstances, and predict mood of a blogosphere.[5]

1.3. Natural Language Processing :

Natural language processing (NLP)[6] is an area of artificial intelligence which allows computers to understand, decode, and regulate human speech. NLP is used to provide an interface between PCs and natural languages. NLP is interested in establishing effective algorithm to process content that is written in a natural language and giving that information reasonably to PC applications. It deals with actual content element processing. The content element is remolded into machine format by NLP. Artificial Intelligence (AI) applies tons of mathematics to determine whether something is negative or positive by utilizing the information which is provided by NLP.

This software package provides automatic extraction of feelings, opinions and sentiments in content and additionally tracks emotions and attitudes on the internet. Individuals express their perspectives by writing blogs, reviews, comments and tweets concerning all types of completely different topics. Tracking brands and items and then determining whether they are viewed negatively or positively will be done by utilizing the internet. Another vital useful case for NLP is sentiment analysis. Information researchers can evaluate comments via web-based networking media to see how their business's label is performing by utilizing sentiment analysis.

2. Literature Survey

This section shows various sentiments analyzing work that has been done in the past. In several sections of this research, the author discusses the significance and uses of sentiment analysis in online communities, as well as the fundamental principles, problems, and a comprehensive investigation [7]. The author of this work outlines in detail the pre-processing methods that must be followed in order to pull out bundle of words from Tweets and offers a subject sentiment analysis methodology. The results of the standard parameter for the subject modeling approach were the focus of the paper. The author of this work proposes an algorithm for turning "bulk of data" from social sites (Twitter) to valuable information and retrieving data by filtering it to meet our needs. Subjects who voice their views commonly have very differing views than others, which is one of the perks of the automatic sentimental analysis given. Most opinions being retrieved in real-time, allowing for faster market reaction times with full - time basis data, making it easy to visualize patterns over time by using R programming language on Tweet. The results of the investigation can be utilized to infer public sentiments, generalize market patterns, and predict outcomes about promising areas. The major goal of this study, according to the authors, was to explain and design a society for Twitter analysis of the data with representation utilizing R and Hadoop, a large data treatment program. We produced a collection of analytic visualization that allows employees to know goods, individuals, products and entertainment data or might obtain insights from this & also took a set of representations, developed in Shiny web apps that aims to combine ui using RHadoop.[8]

3. Working (Data Collection and Cleaning):

3.1. *Sentiment Analysis for Politics awareness:*

Use of sentiment analysis is a method in politics that deals with text information. Such information can be effectively pulled out via web documents to investigate widely reported trends. Politics sentiment analysis identifies regions that are questioned, praised, presented with modifications, or reasoned upon implementation. With more noise in the dataset, the implementation of sentiment analysis algorithms suffers. In this way, dataset cleaning is carried out during pre-processing only with purpose of making it suitable for use with the analysis technique. Because it is an unmonitored method of communication, the online dataset is thought to feature a variety of anomalies and inconsistencies. Tags and advertising that aren't essential to the analyzing process are eliminated. Questionnaire reports that don't show any polarization in sentiment are also taken out. While objective statements in sentiment polar recordings are either eliminated on pre-processing or given to a neutral class in the multi-class categorization with favorable and unfavorable categories. Paragraphs with dual polarity are indeed deleted due to their uncertainty. They do, however, possess polarities and are difficult to assess. Identifying them with the neutral class is also a bad idea, because a favorable and unfavorable attitude does not always imply a neutral sentiment. In reality, elements or viewpoints must be recognized, and feature-level analysis must be undertaken to provide aspect-level outcomes. The type of information isn't retained in unit level thinking, that includes a good, bad, or neutral objective element. This problem is addressed by aspect-based sentiment analysis, which analyses sentiment at the aspect level.[9]

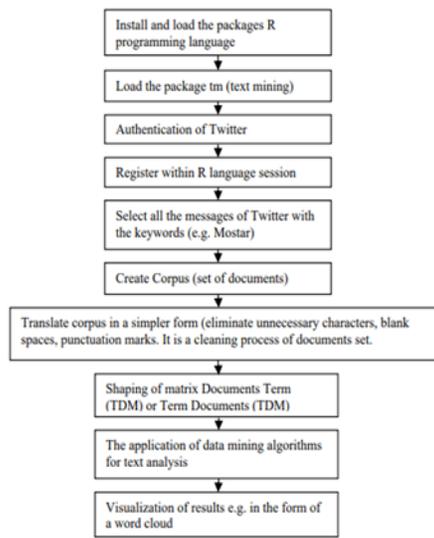


Fig 2 Steps for detecting the views on social network using R Programming Language

3.2. Modules and technologies used

- [10]twitterR is a R program that allows you to use Twitter API. Utility of APIs is endorsed, with a preference for API requests which are more essential for data analysis than in day-to-day contact.
- OAuth is used to provide a user interface to a OAuth 1.0 configuration, allowing users to log to their favored server using OAuth.
- Plyr is a suite of tools for a frequent set of problems: you need to break down a large data structure into uniform parts, assign a function to each component, and then connect all of the results together.
- Dplyr is a reliable and fast tool for dealing with data packets and items, both in and out of memory. Stringr is a collection of wrappers centred on the excellent 'stringi' module, which itself is simple and straightforward when using. The parameter names (and locations) and functions are consistent, all functions handle zero length vectors "NA's" in the same way, and the result of one function may easily be fed into the input of the other.
- Ggplot2 is a framework for making graphics declarative programming, built on "The Grammar of Graphics."
- RcolorBrewer: Gives color schemes for the maps and graphics.
- Sentimentr: At the sentence level, it can estimate text polarity sentiment and optionally aggregate by grouping variable(s) or rows.
- Tm is configuration of text analysis programmes in R.
- Create a word cloud by putting together a list of words.
- Syuzhet: It extracts emotion and attitude plot arcs out of a sentence using a collection of sentiments dictionary that are easily accessible to R users.

3.3. Modules and technologies used

Now, we will make a Twitter application. In this step, we will get a Twitter API which will help us to connect further and gather data or tweets from Twitter directly.

Steps for creating a Twitter application are:

- The very first step is to create your Twitter account.
- After creating an account, go to the site "https://apps.twitter.com" and login to your twitter account.
- After login, go to My Application option → Create a new application.
- Provide all of your information as per the requirements i.e., Application name, description, website name.

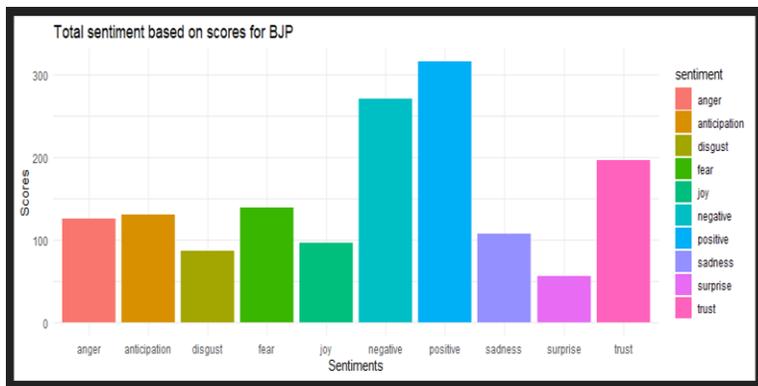


Fig7 Bar Graph for BJP

Table 1. Result for BJP.

Sentiment	Score
Anger	126
Anticipation	131
Disgust	87
Fear	139
Joy	97
Negative	271
Positive	316
Sadness	108
Surprise	56
Trust	197

$$\text{GOOD} = 610(\text{positive} + \text{joy} + \text{trust})$$

$$\text{BAD} = 731(\text{anger} + \text{disgust} + \text{fear} + \text{negative} + \text{sadness})$$

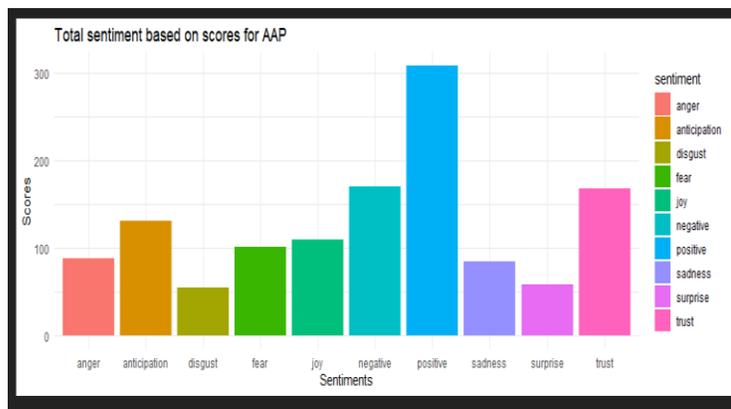


Fig 8 Bar Graph for AAP

Table 2. Result for AAP.

Sentiment	Score
Anger	88
Anticipation	131
Disgust	55
Fear	101

Joy	109
Negative	170
Positive	308
Sadness	84
Surprise	58
Trust	168

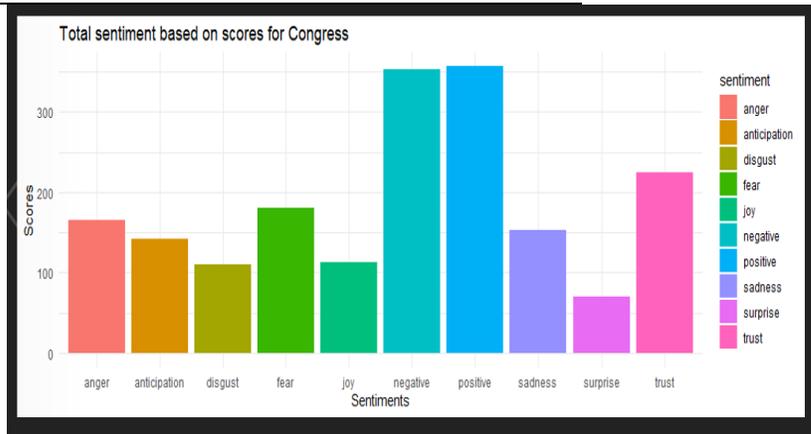


Fig 9. Bar Graph for Congress

Table 3. Result for Congress

Sentiment	Score
Anger	165
Anticipation	142
Disgust	110
Fear	180
Joy	113
Negative	353
Positive	357
Sadness	153
Surprise	70
Trust	225

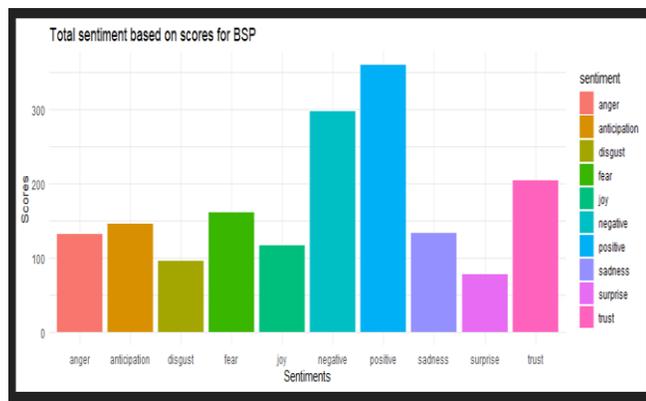


Fig 10 Bar Graph for BSP

Table 4 Result for BSP.

Sentiment	Score
Anger	132
Anticipation	146
Disgust	96
Fear	161
Joy	117
Negative	297
Positive	360
Sadness	133
Surprise	78
Trust	205

5. Conclusions:

By the above given sentiment score and sentiment board we can conclude that for AAP most of the people have positive or good thinking and sentiment about it. Sentiment study on social media for AAP reveals a positive outcome, as the favorable score (585) is greater than its unfavorable score (498). The BJP is in second place.

Table 5. Result for all political parties.

Political Party	Good	Bad
BJP	610	731
AAP	585	498
Congress	595	961
BSP	702	813

References

- [1] P. FICAMOS and Y. LIU, "A Topic based Approach for Sentiment Analysis on Twitter Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 12, pp. 201–205, 2016, doi: 10.14569/ijacsa.2016.071226.
- [2] A. Mittal and S. Patidar, "Sentiment analysis on twitter data: A survey," *ACM Int. Conf. Proceeding Ser.*, vol. 2, no. 1, pp. 91–95, 2019, doi: 10.1145/3348445.3348466.
- [3] S. Saini, R. Punhani, R. Bathla, and V. K. Shukla, "Sentiment Analysis on Twitter Data using R," 2019 *Int. Conf. Autom. Comput. Technol. Manag. ICACTM 2019*, no. April 2020, pp. 68–72, 2019, doi: 10.1109/ICACTM.2019.8776685.
- [4] M. Kamyab, R. Tao, M. H. Mohammadi, and A. Rasool, "Sentiment analysis on Twitter: A text mining approach to the Afghanistan status reviews," *ACM Int. Conf. Proceeding Ser.*, vol. 9, no. 4, pp. 14–19, 2018, doi: 10.1145/3293663.3293687.
- [5] M. T. Khan and S. Khalid, "Sentiment Analysis for Health Care," *Int. J. Priv. Heal. Inf. Manag.*, vol. 3, no. 2, pp. 78–91, 2015, doi: 10.4018/ijphim.2015070105.
- [6] L. M. Nkomo and A. Alm, "Sentiment Analysis," vol. 5, no. 6, pp. 215–231, 2022, doi: 10.4018/978-1-7998-8981-6.ch011.
- [7] P. Khanna, "Available Online at www.ijarcs.info SENTIMENT ANALYSIS : AN APPROACH TO OPINION MINING FROM," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 0976, pp. 252–257, 2017.
- [8] F. Poecze, C. Ebster, and C. Strauss, "Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts," *Procedia Comput. Sci.*, vol. 130, pp. 660–666, 2018, doi: 10.1016/j.procs.2018.04.117.
- [9] E. Younis and E. M. G. Younis, "Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study Spatial and Temporal Environment Impact Analysis on People's Wellbeing View project Big data streaming for disease prediction in Real-time using mac," *Int. J. Comput. Appl.*, vol. 112, no. 5, pp. 975–8887, 2015, [Online]. Available: <https://www.researchgate.net/publication/272463313>
- [10] A. Wankhade, "Social Intelligence Monitoring Based Election Prediction Using Sentimental Analysis in R for Twitter," vol. XIII, no. li, pp. 1057–1065.