IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 3, March 2021

DOI 10.17148/IJARCCE.2021.10312

# Sentimental Analysis in R and Python

**Jasneet Singh Parmar[1], Hritike Singh Rawat[2], Ishaan Nirmal[3]**

Program Manager, De Pedagogics, Gurgaon, Haryana, India[1]

Data Analyst, Tata Consultancy Services, Pune, Maharashtra, India[2]

Post-Graduate, Data Science, Praxis Business School, Kolkata, West Bengal, India[3]

**Abstract**: Twitter sentiment analysis has become very popular because of its usage. Having said that, the stable, optimized Twitter sentiment classification remains elusive due to several issues: heavy class imbalance, domain restrictions of the learning algorithms, the representational richness issues for sentiment cues, and the use of diverse loops. These issues are problematic since many forms of social media analytics rely on accurate underlying Twitter sentiments. Accordingly, a simple yet effective method is proposed for Twitter sentiment analysis. This also does a comparison in both R an python and uses the new code to cover up the existing issues. Experiment results reveal that the proposed approach is more accurate and balanced in its predictions across sentiment classes, as compared to various comparison tools and algorithms. Consequently, this method is better able to reflect strong positive and negative sentiments from users. Considering the importance of Twitter as one of the premier social media platforms, the results have important implications for social media analytics and social intelligence.

**Keywords**: Python, R, Sentimental Analysis, Twitter.

## I. INTRODUCTION

The microblogging site Twitter represents a major platform for users to express their opinions and share their thoughts. Although the "140 characters or less" restriction still applies to every tweet, it has become the very good example of "big data" with a volume and velocity of user contributions that are only paralleled by the no. of input for info about tweets. Taking an example: there are over 2 billion new tweets posted every six days and Twitter receives 30 billion API calls in 2 days (which is almost three times more than any other social platform or search engine). The demand for Twitter data is needed by a growing body of stakeholders interested in a good understanding of what some of the audiences are tweeting about. Due to the fact that most of the tweets are conversational in nature, they often form a very good source for public opinion. What it does is, as a result of its large, different and growing user base, it has emerged as an important source for online opinions & the sentiment indexes that are used by many organizations, governments and individuals. The sentiments polarities can generally be represented as positive, negative, or neutral. Indexes are simply averaged or aggregated sentimental polarity scores that can be used for analysis. This has become an interesting and enlightening topic in which deriving social signals from the Twitter and the similar forms of the social media are now the focus and scrutinization of a plethora of research studies where analysts, scientists are trying to infer public mood about different social events or cultural.

Hence, a robust sentiment classificational analysis performance, the skill to distinguish positive, negative, and neutral sentiments, is vital and useful in many ways. However, it is also known that a stable Twitter sentiment classification performance remains an issue of concern due to a great class imbalance in a problem and representational richness issues for the sentiment cues. These issues are a matter of big concern since many of the forms of media analytics rely on accurate underlying Twitter sentiments. Accordingly, in this study of analysis in R and Twitter, we propose a very simple, optimised framework for Twitter sentiment analysis. The framework uses a "data frame" instead of a conventional "list" and counters the emoji issue which is often used as sarcasm nowadays. The Experiment results reveal that the proposed approach is better and more balanced in its predictions across sentiment classes, as compared to various tools used for comparisons and other ML algorithms. Consequently, this fast optimised sentiment analysis is better able to reflect events eliciting positive and negative sentiments from the users.

Now, Considering the importance of Twitter as one of the most premiere used social media platforms, the outputs have an important implication for social media analytics and social intelligence. It is known that sentiment analysis is often conducted at different levels varying from coarse level to a very fine level. What a Coarse level sentiment analysis deals and does it deal with determining the very sentiment of the full document and the Fine level only deals with the attribute level sentiment analysis. Sentence level sentiment analysis comes in between of these two.
In past and present, many types of research on the area of sentiment analysis of user reviews have been done, performed and utilized. Previous researches show that the performances of sentiment classifiers are dependent on the

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 3, March 2021

DOI 10.17148/IJARCCE.2021.10312

topics. Because of that, we can never say that one individual classifier is the ideal for all of the topics since one classifier does not consistently outperform the other one. Now, we know that the Sentiment Analysis in twitter is quite significantly difficult due to its small and short length. Presence of the emoticons, slang words and misspellings in tweets forced to have a pre-processing step before feature extraction.
There are different feature extraction methods for the collection of the relevant features from the text which can be applied to the tweets also. But the feature extraction is to be done in two phases to extract relevant features. In the first phase, twitter specific features are extracted. Then these features are removed from the tweets to create normal text. Since there is no standard dataset available for the twitter posts of the e-devices, we have taken a random example and real-time tweet. By the doing of sentiment analysis on a particular domain, it is definitely possible to identify the influence of the domain information.

Sentiment analysis – otherwise known as opinion mining – is a much bandied about but often misunderstood term.
In essence, it is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention. Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organisations across the world.

Shifts in sentiment on social media have been shown to correlate with shifts in the stock market. The Obama administration used sentiment analysis to gauge public opinion to policy announcements and campaign messages ahead of 2012 presidential election. Being able to quickly see the sentiment behind everything from forum posts to news articles means being better able to strategies and plan for the future.

It can also be an essential part of your market research and customer service approach. Not only can you see what people think of your own products or services, you can see what they think about your competitors too. The overall customer experience of your users can be revealed quickly with sentiment analysis, but it can get far more granular too.
The ability to quickly understand consumer attitudes and react accordingly is something that Expedia Canada took advantage of when they noticed that there was a steady increase in negative feedback to the music used in one of their television adverts.

**Contextual understanding and tone**
But that is not to say that sentiment analysis is a perfect science at all. The human language is complex. Teaching a machine to analyse the various grammatical nuances, cultural variations, slang and misspellings that occur in online mentions is a difficult process. Teaching a machine to understand how context can affect tone is even more difficult.
Humans are fairly intuitive when it comes to interpreting the tone of a piece of writing.

Consider the following sentence: “My flight’s been delayed. Brilliant!” Most humans would be able to quickly interpret that the person was being sarcastic. We know that for most people having a delayed flight is not a good experience (unless there’s a free bar as recompense involved). By applying this contextual understanding to the sentence, we can easily identify the sentiment as negative.

Without contextual understanding, a machine looking at the sentence above might see the word “brilliant” and categorise it as positive.

## II. LITERATURE SURVEY & ARCHITECTURE

Like the most social media analytics platforms, Twitter harbours a lot of noise, including spam, the short character limit communication style adopted by the users, irrelevant content, and an abundance of the neutral content. All this make Twitter analysis sentiment a very complex task. Not surprisingly, existing tools fall short on delivering a good performance. We here will here compare both R and Python and will ignore using of the learning techniques as they are focused towards only 1 domain. Sentiment Analysis, also called opinion mining or emotion AI, is the process of determining whether a piece of writing is positive, negative, or neutral. A common use case for this technology is to discover how people feel about a particular topic. Sentiment analysis is widely applied to reviews and social media for a variety of applications.

Sentiment analysis can be performed in many different ways. Many brands and marketers use keyword-based tools that classify data (i.e., social, news, review, blog, etc.) as positive/negative/neutral.
Automated sentiment tagging is usually achieved through word lists. For example, mentions of ‘hate’ would be tagged negatively.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 3, March 2021

DOI 10.17148/IJARCCE.2021.10312

There can be two approaches to sentiment analysis:

1. Lexicon-based methods
2. Machine Learning-based methods.

In this research, we will be using a Lexicon-based method. Lexicon based methods define a list of positive and negative words, with a valence—(eg 'nice': +2, 'good': +1, 'terrible': -1.5 etc). The algorithm looks up a text to find all known words. It then combines their individual results by summing or averaging. Some extensions can check some grammatical rules, like negation or sentiment modifier (like the word "but", which weights sentiment values in text differently, to emphasize the end of text).

In the architecture below, we will first do the requisite authentication, followed by the second process that is to do the extraction of the tweets from where all the data will come and inevitably the analysis.

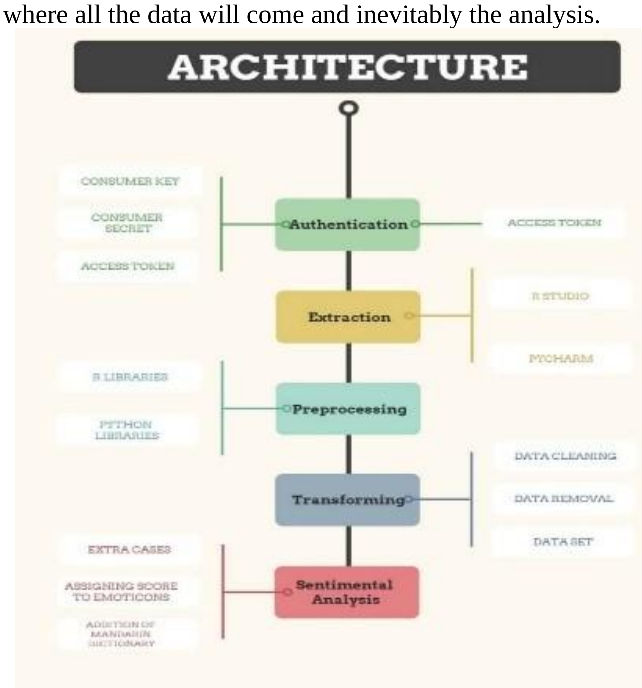


Fig 1. Architecture of Sentimental analysis

## III.METHODS

**Twitter Sentiment Analysis using R**
The Implementation of the sentiment analysis application in R is as follows:

1. Extraction of the tweets using the Twitter app
2. Proper processing of the Clean tweets which are used for the further analysis
3. Get the requisite sentiment score for an individual tweet
4. Segregation of the positive and the negative tweets

**Twitter Sentiment Analysis using Python**
In this process, we actually determine whether the tweets which people used to write to show their views towards any points, brands, company etc. are actually on the positive side or on the negative side.
In python, we can do the twitter analysis with the help of mainly 2 modules or packages of python. These are:

1. Textblob: This module or library helps the programmer to tokenize the tweets which means it helps the coder to splits the word from the full-length text or sentence so that it becomes easier for the programmer to analyse the tweet into positive, negative or neutral. The text blob is also used to remove the irrelevant words from the tweets which generally creates a lot of problem during the analysis of the tweets. For ex-(I, am, is, are, the etc.) Textblob is also used to select only significant or meaningful words from the text to analyse i.e it always takes only adjectives or adverbs for the analysis and ignore the other unnecessary text or sentence. It also helps in passing the tokens or necessary words which is used for analysis to a special type of classifier generally known as sentiment classifier which helps the coder

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 3, March 2021

DOI 10.17148/IJARCCE.2021.10312

to classify the particular tweets into negative, positive or neutral. If the polarity score comes between 0 to -1, then the tweet is considered as negative and if the polarity score comes between 0 to 1, then the tweets are considered as positive.

2. Tweepy: It is also a part of the python libraries which is generally used to get the access to the python to communicate with the Twitter platform and use its API which is generally termed as Application Program Interface and is basically a type of protocol which usually specifies how software can interact. It is used for making software applications. Tweepy is also very much important for the analysis of the tweets as it helps the programmer to connect the python with its Twitter account.

## IV. RESULT & ANALYSIS

**Sentiment Analysis Issues to Be Aware:**

1. Sarcasm: is one of the most difficult sentiments for automated tracking to interpret properly. Example: “It was awesome for the week that it worked.”
2. Navel gazing: is when social media tracking turns up items related to your own promotional efforts, and should be filtered out.
3. Neutral sentiment: is similar to the concept of swing voters, and Frank recommended dividing it into specific themes to uncover more detailed opinions.
4. Relative sentiment: is not a classic negative, but can be a negative nonetheless. Example: “I bought an iPhone” is good for Apple, but not for Nokia.
5. Compound or multidimensional sentiment: contain positives and negatives in the same phrase. Example: “I love Mad Men, but hate the misleading episode trailers.”
6. Conditional sentiment: includes actions that may happen in the future. Example: the customer isn’t angry now but says he will be if the company doesn’t call him back.
7. Positive feelings can be unrelated to the core issue. For example, many comments about actors focus on their personal lives, not their acting skills.
8. Negative sentiment is not necessarily bad: This relates to the classic PR dilemma regarding negative publicity. Example: Sarah Palin’s appearance on the Today show generated many negative comments but still drove ratings increases.
9. Ambiguous negative words: Their context needs to be thoroughly understood and tagged accordingly. Example: “That backflip was so sick” is really a positive statement.
10. Beware of Google translate syndrome: Western and Asian sentiment differ greatly, as do the meanings of their emoticons, so they need to be interpreted correctly.

## V. CONCLUSION

In this study, we proposed and evaluated effectively a program for Twitter sentiments. We did a comparison in both R and Python by implementing both of them. Experiment results revealed that R was able to provide an easier way to provide more details because of the easy conversion of the list into a data frame. A higher detailed explanation of the tweets as compared to various comparison methods evaluated in Python. When the comparisons were done on different parameters, the further analysis revealed that the Polarity in Python gives higher performance than in R since it is more near to the actual value and is not using whole number polarity. As a result, in Python, it produced more accurate and balanced representations of sentiment polarity indexes in social media but R provided more functionality. Given the extensible nature of this research, future work can expound upon it by extending the expansion parameter values (data sets, feature sets, and classifiers) and by improving on the search method used and the different loops used. Given the importance of Twitter as one of the premier social media platforms, the results of this work have important implications for social media analytics and social intelligence.

## ACKNOWLEDGMENT

**Dr. R.P. Mahapatra** (Prof., HOD CSE, SRM-IST), **Analp Pathak** (Assistant Professor, SRM-IST), **Govind Verma** (Assistant Professor, SRM-IST) are gratefully acknowledged for their inputs and discussions during this study.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 3, March 2021

DOI 10.17148/IJARCCE.2021.10312

## REFERENCES

[1]. Crews.T.B. & Stit-Godes WL (2012). Incorporating Facebook and Twitter la service learning project in a business communication course. Business Communication Quarterly, 7517679

[2]. Cochrane, T. (2016). Twitter tales Facilitating international collaboration with mobile web 2.0. In C.H.Steel,M.J. Keppell, P. Gerbie S. Howego (Eds.) Curriculum, technology transformation for an unknown future.

[3]. Proceedings ascite Sydney 2010 (pp. 195-205) Tan, J.PL. (2008). Teaching smarter: How mining ICT data can inform and improve learning and teaching practice. In Hello Where are you in the landscape of educational technology.

[4]. Proceedings ascilite Melbourne 2008 http://www.scilite.org.au/conferences/melbourne00/proca dawson Dekker, AH. (2011) Temporal social network analysis of discourse, MODSIM 2011, 19th International Congress an pp. 447-453 Kassens-Noor, E (2012)

[5]. Twitter as a teaching practice to enhance active and informal learning in higher education: The case of sustainable tweets Active Learning in Higher Education, 13111921 Lowe, & affey, D.(2011) Is Twitter for the hindising Twitter to enhance student learning in a marketing course Journal of Marketing Education, 332183192 Phillips, R., Maor, D, Cumming-Potwin, Roberts, P Herrington 2. Preston, G.& Moore, E. (2011)

[6]. Learning analytics and study behaviour A pilot study. In G. Williams, P.Statham Brown&B Cleland (Eds. Changing Demands, Changing Directions Proceedings site Hobart 2011. (pp.997)

## BIOGRAPHY

**Jasneet Singh Parmar**, B. Tech in Computer Science and Engineering (2015-2019) from SRM Institute of Science and Technology, Chennai, India. Currently working as Program Manager at De Pedagogics, Gurgaon, Haryana, India.

**Hritike Singh Rawat**, B. Tech in Computer Science and Engineering (2015-2019) from SRM Institute of Science and Technology, Chennai, India. Currently working as Data Analyst at Tata Consultancy Services, Pune, Maharashtra, India.

**Ishaan Nirmal**, B. Tech in Computer Science and Engineering (2015-2019) from SRM Institute of Science and Technology, Chennai, India. Currently pursuing Post-Graduation in Data Science from Praxis Business School, Kolkata, West Bengal, India