



# Brand Perception Using Social Media Analysis

**Aksheya Suresh<sup>1</sup>; Dinesh Kumar P<sup>\*2</sup>; Azar Mahmoodh Y<sup>\*3</sup>; Delliganesh S<sup>\*4</sup>**

<sup>1</sup>Assistant Professor, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India

<sup>2,3,4</sup>Student, Rajalakshmi Engineering College, Computer Science and Engineering, Chennai, India

Email: [aksheya.suresh@rajalakshmi.edu.in](mailto:aksheya.suresh@rajalakshmi.edu.in)<sup>1</sup>, [dineshdk1898@gmail.com](mailto:dineshdk1898@gmail.com)<sup>2</sup>, [azarmahmoodh021998@gmail.com](mailto:azarmahmoodh021998@gmail.com)<sup>3</sup>, [dilliganesh1806@gmail.com](mailto:dilliganesh1806@gmail.com)<sup>4</sup>

**Abstract**— In recent years, the interest among Microblogging has become a very popular communication tool among Internet users and a way to monitor the public's feelings towards their brand, business, products etc. Social media especially twitter is providing a space for expressions and opinions, where users discuss various events, services, and brands. Organizations are in continuous need of the feedback about their services to improve the quality and quantity. However, due to the bulk amount of data, it's difficult to detect the consumer's opinions. The ability to process this information has become important to deep dive into the Brand Perception using social media analysis. Brand perception is a special result of a consumer's experiences with a brand. The main purpose of this project is to find the public opinion about the international brands among the consumers, what they really think and feel about their brand.

**Keywords**— Polarity, Sentiment, Opinion, Analysis, Twitter.

## Introduction

The emergence of social media has given internet users a venue for expressing and sharing their thoughts and opinions on all types of topics and events. Twitter, with nearly 600 million users and over 250 million messages per day, has quickly become a gold mine for organizations to observe their name and types by extracting and analyzing the sentiment of the tweets announce by the general public concerning them, their markets, and competitors [2]. Microblogging platforms square measure utilized by totally {different |completely different} folks to specific their opinion concerning different topics, so it's a valuable supply of people's opinions. As additional and additional users post concerning product and services they use, or categorical their political and non-secular views, microblogging web-sites become valuable sources of people's opinions and sentiments [1].

We collected a corpus of text posts from Twitter equally split mechanically between 3 sets of texts:

1. Texts containing positive emotions, like happiness, amusement or joy.
2. Texts containing negative emotions, like disappointment, anger or disappointment.
3. Objective texts that solely state a reality or don't categorical any emotions [2].

The task of sentiment analysis is troublesome thanks to the variability and quality of language expressions. One amongst the most important challenges of this task is led to by defeated and negated expressions. A defeated expression contains variety of words that have a polarity that is opposite to the polarity of the expression itself.

**Related Works**

1: With the population of blogs and social networks, opinion mining and sentiment analysis became a field of interest for several researches. a really broad summary of the prevailing work was given in (Pang and Lee, 2008) [2]. In their survey, the authors describe existing techniques associated approaches for an opinion-oriented data retrieval. However, not several researches in opinion mining thought of blogs and even abundant less self-addressed microblogging. In (Yang et al., 2007), the authors use web-blogs to construct a “Corpora” for sentiment analysis and use feeling icons assigned to journal posts as indicators of users’ mood. The authors applied SVM and CRF learners to classify sentiments at the sentence level and so investigated many methods to see the general sentiment of the document. because the result, the winning strategy is outlined by considering the sentiment of last statement.

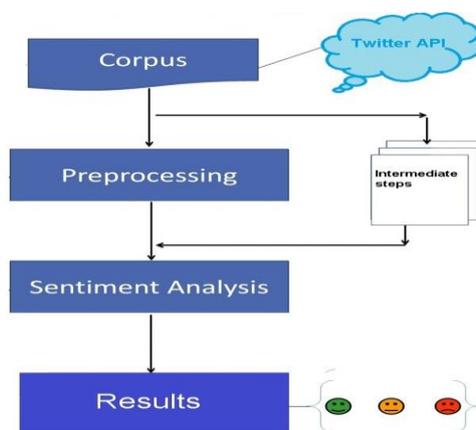
2: Agarwal et al additionally explored the POS options, the lexicon options excluding merely combining numerous options, they additionally designed a tree illustration of tweets to mix several classes of options in one compendious illustration [4]. A partial tree kernel was accustomed calculate the similarity between 2 trees. They found that the foremost vital options square measure people who mix previous polarity of words with their POS tags. All different options solely play a marginal role. moreover, they additionally showed that combining unigrams with the simplest set of options outperforms the tree kernel-based model and offers concerning four-dimensional absolute gain over a unigram baseline.

3: Existing work in the main concentrates on the utilization of 3 sorts of options; lexicon features, POS options, and microblogging options for sentiment analysis. Mixed findings are according. Some argued the importance of POS tags with not word previous polarity concerned, whereas others emphasized the utilization of microblogging options [6]. during this paper, we have a tendency to propose a replacement form of options for sentiment analysis, referred to as linguistics options, wherever for every entity in an exceedingly tweet (e.g. iPhone, iPad, MacBook), the abstract idea that represents it’ll be further as a replacement feature (e.g. Apple product). We have a tendency to compare the accuracy of sentiment analysis against different sorts of features; unigrams, POS options, and also the sentiment-topic options. To the simplest of our data, mistreatment such linguistics options is novel within the context of sentiment analysis.

**System Design**

Using Twitter API we have a tendency to collect a corpus of text posts and fashioned a dataset of 3 classes: positive sentiments, negative sentiments, and a collection of objective texts to gather negative and positive sentiments, we have a tendency to followed an equivalent procedure as in (Read, 2005; Go et al., 2009). We have a tendency to queried Twitter for 2 sorts of emotions:

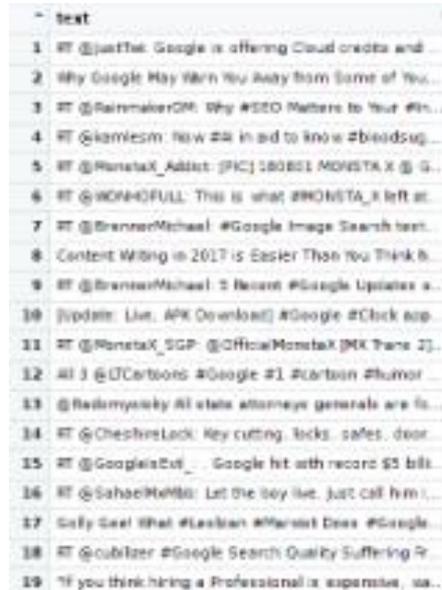
- Happy emoticons: “:-)”, “:)”, “=)”, “:D” etc.
- unhappy emoticons: “:-(”, “:(”, “=(”, “;(” etc.



**Fig. 1** Architecture Diagram

## 1. Data Formation

Data Formation could be a document, it contains twitters supported the keywords provided. In our research, we use English language. However, our method will be adapted easily to other languages since Twitter API allows to specify the language of the retrieved posts. Within the graph, we see that superlative adjectives are used more often for expressing emotions and opinions, and comparative adjectives are used for stating facts and providing information.



**Fig. 2** Formation Of Text Data

## 2. Data Preprocessing

Data preprocessing is done to eliminate the incomplete, noisy and inconsistent data. Data must be preprocessed in order to perform any data mining functionality. Data Preprocessing involves the following tasks.

**2.1: Removing URLs:** Normally URLs doesn't contribute to investigate the sentiment within the informal text. For instance, consider the sentence “I have logged in to [www.Ecstasy.com](http://www.Ecstasy.com) as I’m bored” actually the above sentence is negative but due to the presence of the word ecstasy it should become neutral and it’s a false prediction. So as to avoid this type of failures we must employ a method to get rid of URLs.

**2.2: Filtering:** Usually people use repeated letters in words like happyyyy to indicate their intensity of expression [5]. But, these words don't seem to be present within the sentiwordnet hence the additional letters within the word must be eliminated. This elimination follows the rule that a letter can’t repeat quite 3 times hence can eliminate such letter.

**2.3: Removing Special Characters:** Special characters like.,[]()/? should be removed so as to get rid of discrepancies during the assignment of polarity. For instance, “it’s good:” if the special characters don't seem to be removed sometimes the special characters may concatenate with the words and make those words unavailable within the dictionary [5]. so as to beat this we remove special characters.

**2.4: Removal of Retweets:** Retweeting is the process of copying another user's tweet and posting to a different account [1]. This happens if a user likes another user's tweet. Retweets are commonly abbreviated with \RT."



#### 4. Clustering

K-Means clustering may be a popular clustering algorithm supported the partition of knowledge [4]. Data that have the same characteristics are grouped into one cluster, whereas data that have different characteristics are grouped into other clusters. Steps for K-Means clustering are as follows,

a. Decide the number of cluster K.

b. Initialization of the cluster center (centroid). It can be conducted by using various ways [6]. However, the most frequent way is by using random way. Clusters centers are assigned by random numbers.

c. Allocate all data/objects to the closest cluster. Determination of closeness of two objects is determined based on the distance of two objects [7]. For calculating the distance of all data to each centroid point, Euclidean Distance theory is used, which is formulated as follows.

$$D(i, j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ip} - X_{jp})^2}$$

Where:

- $D(i, j)$  = distance of  $i^{th}$  data to cluster center  $j$
- $X_{ik}$  =  $i^{th}$  on the  $k^{th}$  data attribute
- $X_{ij}$  =  $j^{th}$  center point on the  $k^{th}$  data attribute

d. Recalculate centroid with current cluster membership. Centroid is an average (mean) of all data/objects within particular cluster. If desired, the median of this cluster can also be used.

e. Reassign each object by using new cluster centre, if the cluster doesn't change, then clustering process finished otherwise repeat step 3 until there is no change for each cluster.

#### Conclusion

This study has developed various novelties. Among them, in this study, K-Means which has generated a better accuracy when it was combined with K-NN like the research in related works [1]. It was done to get the simplest data accuracy. In our research, we've presented a way for an automatic collection of a corpus which will be used for further analysis. We used the collected corpus to coach a sentiment classifier [3]. Our classifier is in a position to work out positive, negative and neutral sentiments of documents. Sentiment Analysis has cause development of higher products and good business management. This research area has provided more importance to the mass opinion rather than word-of-mouth. In addition, we could take the time to gather far more data (a factor of 10 times more, perhaps), ensuring to implement the right preprocessing techniques to make sure that overfitting doesn't occur in the widely varying class of language that can occur on Twitter. The explosive growth of Twitter has attracted considerable media and consumer attention to the present service. In terms of long-term development, the last word utility of our sentiment analyzer is to create a classifier adapt enough to mine the Twitter database given a particular keyword, Text Processing and Sentiment analysis emerges as a challenging field with many obstacles because it involves text processing. It has a good sort of applications that would collect its results, like news analytics, marketing, question answering, readers do. Getting important insights from opinions expressed on the web especially from social media blogs is significant for several companies and institutions, whether it's in terms of product feedback, public mood, or investors opinions. Sentiment analysis is a difficult technology to get right [5]. However, when you do, the benefits are great. Look for a tool that has uses text Processing technology and ideally with machine learning capabilities. Look for a vendor that treats sentiment analysis seriously and shows advancements and updates in their sentiment analysis technology.

## References

1. Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", Universit e de Paris-Sud, Laboratoire LIMSI-CNRS, B atiment 508, F-91405 Orsay Cedex, France.
2. Ravi Parikh and Matin Movassate, "Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques".
3. Muhammad Ihsan Zul, Feoni Yulia, Dini Nurmalasari, "Social Media Sentiment Analysis Using K-Means", Informatics Engineering Politeknik Caltex Riau Pekanbaru, Indonesia.
4. Namrata Godbole, Manjunath Srinivasaiah, Steven Skiena, "Large-Scale Sentiment Analysis for News and Blogs", Dept. of Computer Science, Stony Brook University, Stony Brook, NY 11794-4400, USA.
5. Cacia Zirn, Mathias Niepert, Heiner Stuckenschmidt, Michael Strube, "Fine-Grained Sentiment Analysis with Structural Features", KR & KM Research Group University of Mannheim Mannheim, Germany.
6. Amandeep Kaur, Vishal Gupta, "A Survey on Sentiment Analysis and Opinion Mining Techniques", University Institute of Engineering and Technology, Chandigarh, India.
7. Ray Chen, Marius Lazer, "Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement".