

# Sentiment Analysis of Tweets for Inferring Popularity of Mobile Phones

Hema Krishnan  
Research Scholar  
School of Engineering  
CUSAT, Kochi -21

M. Sudheep Elayidom  
Associate Professor  
School of Engineering  
CUSAT, Kochi -21

T. Santhanakrishnan  
Scientist E  
NPOL  
Thrikkakara, Kochi-21

## ABSTRACT

Sentiment analysis is a very popular technique for social network analysis. Sentiment analysis also termed as opinion mining is a process of automatically extracting knowledge from sentiments or opinions of others about some topic or problem. We can identify opinions in a large unstructured/structured data and analyze the polarity of opinions. Twitter is a large and rapidly growing micro blogging social networking website where people express their opinions in a short and simple manner of expressions. It is a common practice that merchants selling products on the Web ask their customers to review the products. In twitter number of customer reviews on different products is appearing. Mobile phones are a common domain in which number of customer reviews appears. This makes it difficult for a potential customer to read them in order to make a decision on whether to buy the product. We are only interested in the specific features of the phones that customers have opinions on and also whether the opinions are positive or negative. This paper presents a lexicon based approach for analyzing the customer reviews on mobile phones over Twitter data to measure the popularity based on which the customer can decide whether to buy the product.

## Keywords

Sentiment, opinion, score, Twitter, lexicon

## 1. INTRODUCTION

Most of the companies today improvise their products using different means. Reaching out to customers is a common means adopted by companies to understand the customer sentiments. Customer feedbacks are usually obtained using feedback forms, customer satisfaction surveys, reviews and activity tracking. Based on the data obtained from these feedbacks, the companies can improvise their products and services associated with it. For example: e-commerce websites track the user click patterns and the recommendation engines suggest products accordingly. There are many tools available for sentiment analysis using the large data available on different social networking sites like Facebook, Twitter, LinkedIn etc. Mostly sentiment analysis applications are used in prediction engines. For example: Election results prediction. These predictions are usually done by building a model on the data. Sentiment analysis has been previously implemented in customer feedback system, prediction engines and recommendation engines.

This paper presents a novel lexicon based sentiment analysis technique which is used to analyse the sentiment of people on different mobile phones. The user selects the input on the mobile phones and the brand of the device via radio buttons, and drop down menus respectively. Based on the input, graph is created showing the sentiment score distribution of each mobile phone brand.

## 2. RELATED WORKS

A lot of work has been done in the field of opinion mining or sentiment analysis for well over a decade now. Twitter is different to other forms of raw data which are used for sentiment analysis as sentiments are conveyed in one or two sentence blurbs rather than paragraphs. There are various approaches for mining twitter data. A system that processes the tweets by pulling data from tweeter posts was developed by Das et al. [1]. Data collected from twitter were preprocessed and connected to Alchemy API. Unstructured contents (news, articles, blogs, posts etc.) can be analyzed by the web service Alchemy API.

Twitter is much more informal and less consistent in terms of language. Users cover a wide array of topics which interest them and use many symbols such as emoticons to express their views on many aspects of their life [2]. When using human generated status updates, sentiment are not always obvious; many tweets are ambiguous and can use humour to maximize the opinion to other human readers but deflect the opinion to a machine learning algorithm [2]. Another consideration when using a dataset generated from Twitter is that a considerably large amount of tweets which convey no sentiment such as linking to a news article, which can lead to difficulties in data gathering, training and testing [3]. Sentiment analysis provides a means of tracking opinions and attitudes on the web and determines if they are positively or negatively received by the public. Different techniques are used to classify the text according to polarity. Most of these techniques can be classified under two categories.

### 2.1 Machine Learning

Machine learning strategies [4] work by training our algorithm with a training data set before applying it to the actual data set. That is, a machine learning algorithm needs to be trained first for both supervised and unsupervised learning tasks.

### 2.2 Lexicon Based

Lexicon Based techniques work on an assumption that the collective polarity of a document or sentence is the sum of polarities of the individual words or phrases.

The proposed approach is a lexicon based technique i.e. a dictionary of sentiment bearing words along with their polarities was used to classify the text into positive, negative or neutral opinion. Machine learning techniques are not used because although they are more accurate than the lexicon based approaches, they take far too much time performing sentiment analysis as they have to be trained first and hence are not efficient in handling big sentiment data.

### 3. NOVEL APPROACH FOR SENTIMENT ANALYSIS

Our novel approach aims to analyze, design and implement sentiment analysis over Twitter data of Mobile Phones and to infer phone Popularity.

The basic steps of the lexicon based technique [4] are outlined below:

- Preprocess each text (i.e. remove HTML tags, noisy characters).
- Initialize the total text sentiment score:  $s = 0$ .
- Tokenize text. For each token, check if it is present in a sentiment dictionary.

– If token is present in dictionary,

- If token is positive, then  $s = s + 1$ .
- If token is negative, then  $s = s - 1$ .

• Look at total text sentiment score  $s$ ,

- If  $s > \text{threshold}$ , then classify the text as positive.
- If  $s < \text{threshold}$ , then classify the text as negative.

The proposed methodology can be divided into different modules.

#### 3.1 Fetching Tweets

In this module, tweets are fetched from the twitter website. The package `twitter` is imported for the same. There are two sub modules within this module.

##### 3.1.1 Twitter OAuth

Here the application should authenticate with twitter so as to access the twitter data. For this, an account has to be created in twitter. Four parameters [5] will be provided by twitter for the developers using which the developers and their applications can be authenticated. The four parameters are: consumer key, consumer secret, access token and access secret. After authentication, data can be fetched from twitter.

##### 3.1.2 Extraction of Tweets

Tweets are extracted from the site using `searchTwitter()` function. The function takes the following inputs:

- Search query to be issued to twitter.
- Maximum number of tweets to be returned.
- Language.
- Date since which the tweets have to be obtained.

The search for tweets is done using the hashtag of respective phone name. The search can also be done using other twitter handles as well.

#### 3.2 Processing and Storing Tweets

In this module, tweets after processing are stored into the database. `Mongodb` is used for the database. This module can be viewed as two submodules.

##### 3.2.1 Processing of Tweets

Here the tweets are subjected to a set of procedures which makes them suitable for sentiment analysis in the next module. The following steps are involved in the processing of tweets:

- Removal of re-tweets.
- Removal of handles like # etc.
- Removal of punctuations.
- Removal of digits.
- Removal of web addresses like `https://` etc.
- Removal of blank spaces.
- Convert the entire text to lower case.

##### 3.2.2 Storing Tweets

The processed tweets are then converted into a dataframe and inserted into a `mongodb` [6] collection. `MongoDB` database is a NoSQL document oriented database which stores the data in JSON format.

Lexicon based approach is used to understand the sentiments of the tweets. Each tweet will have a sentiment score associated with it. The distributions of sentiment scores for different tweets are plotted in a graph. Each brand has a separate graph for it. A dictionary containing the possible adjectives/words indicating emotions is available. Each word will have a score associated with it. The scores range from  $-5$  to  $+5$ . A negative score indicates a negative sentiment and a positive score indicates a positive sentiment. Five classes of sentiments and their score range is given below:

- very positive  $+3$  to  $+5$ .
- positive  $0$  to  $+3$ .
- neutral  $0$ .
- negative  $-1$  to  $-3$ .
- very negative  $-3$  to  $-5$ .

### 4. IMPLEMENTATION AND RESULTS

The lexicon based approach is implemented using R and `MongoDB`. Five mobile phone brands are considered for study- Samsung, nexus, iphone, lenovo, motorola. Main packages used in R are `twitter`, `mongolite`, `ggplot2`, `ggmap` and `shiny`. Interface is created using `shiny`. Latest 500 tweets related to each phone brand is extracted. Graph is plotted for five mobile phones showing the sentiment scores versus number of tweets related to each brand.

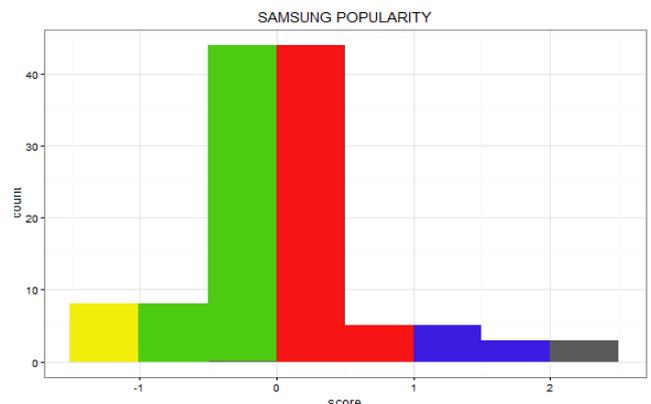
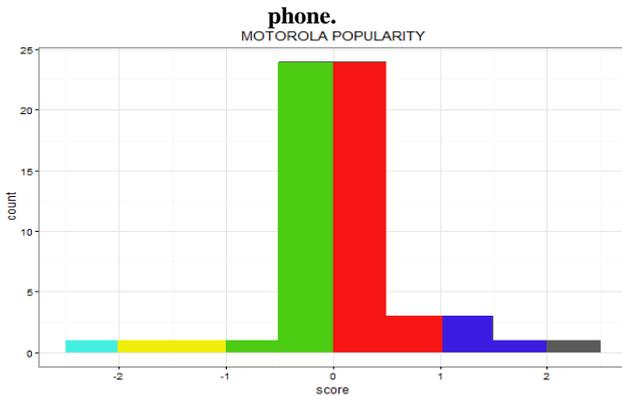
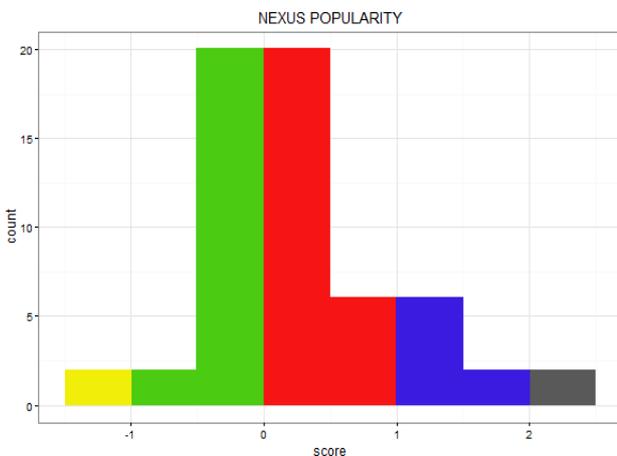


Fig. 1. Plot showing the sentiment score for samsung



**Fig. 2. Plot showing the sentiment score for motorola phone.**



**Fig. 3. Plot showing the sentiment score for nexus phone**

Figs. 1–3 shows the various plots for Samsung, Motorola and Nexus phone brands. From the above plot it can be inferred that the various Samsung phone tweets have sentiment scores between  $-1$  and  $+2$ . We can see the scores vary between different values. Different range of scores shows the sentiments of the people regarding the phone. Similarly the

scores range between  $-2$  and  $+2$  and between  $-1$  and  $+2$  respectively for Motorola and Nexus brands respectively. In each plot the popularity of phone can be inferred by analysing the sentiment scores. Customers may be able to make a decision to buy a new phone brand by analysing the popularity.

## 5. CONCLUSION

Sentiment analysis is a technique for analyzing the product popularity. In this paper we have considered five mobile phone brands and inferred their popularity by analyzing their sentiment scores. We have used a lexicon based approach for assessing popularity. To improve accuracy, in future we can use hybrid approaches.

## 6. REFERENCES

- [1] T.K. Das, D.P. Acharjya, M.R. Patra, "Opinion Mining about a Product by analysing public tweets in Twitter", IEEE Proceedings of International Conference on Computer Communication and Informatics (ICCI-2014), January 03–05, 2014, Coimbatore, India.
- [2] Varsha Sahayak, Vijaya Shete, Apashabi Pathan, "Sentiment Analysis on Twitter Data", International Journal of Innovative Research in Advanced Engineering (IJIRAE), Issue 1, Volume 2, January 2015.
- [3] Hema Krishnan, M. Sudheep Elayidom, T. Santhanakrishnan, "Impact and Application of Sentiment Analysis using Twitter: A Survey", International J. of Advanced Research in Computer and Communication Engineering, (IJARCCE), Vol. 4, Special Issue 1, 2015, pp. 18–21.
- [4] S.M. Vohra, J.B. Teraiyam "A Comparative Sentiment Analysis Technique", Journal Of Information, Knowledge And Research In Computer Engineering, ISSN: 0975 – 6760, Vol. 2, Issue 02, 2013.
- [5] <https://sites.google.com/site/miningtwitter/questions/sentiment/>.
- [6] <http://www.tutorialspoint.com/mongodb/>.