

## Twitter Sentiment Polarity Classification & Feature Extraction

Mrs. Smita Bhanap<sup>1</sup>, Dr. Seema Kawthekar<sup>2</sup>

<sup>1</sup>(Fergusson College, Pune, India)

<sup>2</sup>(Department of CSIT, Dr. B.a.M.U. Aurangabad, India)

---

**Abstract:** *Advancements in social networking site and usage by people for various purposes like posting reviews of products which they purchase online has gained attention of researchers. These reviews are real time messages about people's opinions on different topics, current issues, complaints about products. One of the most interesting and challenging issue is to analyze these reviews on social sites such as Twitter in order to get valuable hidden patterns. In this paper, we have focused on classifying polarity of tweets into either positive, negative or neutral tweets. Also, we discussed how machine learning algorithms can be applied for classification and feature selection. Future enhancement focuses on how these customer reviews plays a key role in building business decision support systems. Experimental evaluations show how to classify tweets of electronic product datasets.*

**Keywords:** *twitter datasets, polarity classification, feature extraction, DSS.*

---

### I. Introduction

Microblogging today has turned into an exceptionally well known specialized service among Internet clients. Every day, millions of users share opinions on various aspects of life. Hence microblogging sites contain huge data for opinion mining and sentiment analysis. Authors of those messages write about their life, share opinions on variety of topics and discuss current issues. With the quick development of Web 2.0, users generate tremendous quantities of online information in the forms of reviews, blogs, tweets, etc. People share their opinion regarding a wide range of products, their features and services which help business people in decision making. Sentiment analysis also referred to as Opinion mining, is the field of study that analyses people's opinion, sentiments, attitudes, evaluations, and emotions through social media data. These topics are most likely to be covered by reviews. Sentiment analysis in reviews is the process of investigating product reviews on the internet to determine the feeling or overall opinion about a product [2].

### II. Motivation And Objective

Research work in this area shows that a number of approaches are used to identify the significant features of opinion mining and to determine the sentiment of the text, whether it is positive or negative, which is extended to strength of polarity. Polarity of tweets helps in understanding views of customer about a product. These reviews helps other customers in deciding before purchasing a product and also business organizations in decision making for better quality production. In this paper we analyze the tweets downloaded using twitter API and find its polarity. These polarity scores can be further used to classify various features of customer reviews.

### III. Related Work

Authors of [1] has introduced Sentiment Analysis in Twitter. The opinions and reviews collected from Twitter and SMS are classified to sentiment using contextual phrase-level polarity. The sentiments were classified in three ways positive, negative or objective. Sindhu C focused on a Survey on Opinion Mining and Sentiment Polarity Classification.[1]

For the sentiment classification, author in [5] collected texts containing emoticons from Usenet newsgrouphaving emoticons which are used to create a training set. It categorized the dataset into "positive" and "negative" having the texts with happy emoticons and sad or angry emoticons respectively. Authors make use of SVM and Naive Bayes algorithm to train the dataset, and obtain up to 70% of accuracy on the test set. In [6], authors used Twitter to collect training data and performed a sentiment analysis. Here they developed corpora by utilizing emoticons to acquire "positive" and "negative" samples, and then also utilize different classifiers. The authors got best result with the Naive Bayes classifier using a mutual information measure for feature selection. The authors were able to obtain up to 81% of accuracy on their test set. However, this strategy displayed a terrible performance with three classes "positive", "negative" and "neutral" [7]. In [8], author did sentiment classification on Twitter data by using different syntax features of tweets like retweet, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words [9].

#### IV. Sentiment Analysis

Sentiment analysis, also called Opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes. The sentiment analysis can be performed at one of the three levels: the document level, sentence level, feature level.[2]

**Document Level Sentiment Classification:** In document level sentiment analysis it is required to extract informative text for inferring sentiment of the whole document. It is difficult for learning methods to deal with objective statements which can be rendered by subjective statements. This may lead to complicate further for document categorization task with conflicting sentiment.

**Sentence Level Sentiment Classification:** In this type polarity of the sentence can be given by three categories as positive, negative and neutral. It is a challenging area to deal with identification features indicating whether sentences are on-topic which is kind of co-reference problem.

**Feature Level Sentiment Classification:** Feature level focuses on product features which are defined as product attributes or components. Feature based sentiment analysis means analysis of such features for identifying sentiment. In this approach positive or negative opinion is identified from the already extracted features.

#### V. Data Collection And Preprocessing

Twitter is a social networking service that lets its users to post real time messages, called tweets. Tweets have many unique characteristics. Twitter, with nearly 600 million users and over 250 million messages per day, has rapidly turned into a gold mine for organizations to monitor their reputation and brands by extracting and analyzing the sentiment of the Tweets posted by the public about their remarks, markets, and other contenders. Performing Sentiment Analysis on Twitter is complicated than doing it for large reviews. This is because the tweets are very short and mostly contain slangs, emoticons, hash tags and other twitter language.

**Following steps are followed for preparing sentiment score of tweets.**

**Step 1:** Tweets of various products are downloaded using twitter API.

**Step 2:** Downloaded tweets are cleaned to remove noise. The data containing hashtags, URL, punctuations, special symbols conversion to lowercase is done.

**Step 3:** Cleaned data is then used for sentiment extraction and classification purpose.

**Step 4:** Classification and feature extraction is done. Positive and negative tweets are identified and compared with dictionary of positive and negative words.

#### VI. Sentiment classification and feature selection:

Sentiment Classification can be done using Machine Learning approach and Lexicon Based approach. Machine learning approach uses classification of sentiment using training and test data sets. The text classification methods using ML approach can be differentiated into supervised and unsupervised learning methods [10]. Machine learning techniques like Naive Bayes (NB) and support vector machines (SVM) have made extraordinary success in text categorization.

NB: Naive Bayes classification model works with Bag-of-Word feature extraction which avoids the position of the word in the document. It predicts the probability that a given feature set belongs to a particular label.[4] SVM: SVM can build a nonlinear decision surface in the original feature space by mapping the data instances non-linearly to an inner product space where the classes can be linearly separated with a hyper-plane [11]

Lexicon Based Approach is based on finding the opinion lexicon which is used to analyze the text. Opinion words are utilized in many sentiment classification tasks. Positive opinion words are utilized to express some desired states, while negative opinion words are utilized to express some undesired states [12].

Lexicon based approach uses dictionary and corpus based approach. In dictionary-based approach we find opinion seed words and then searches the dictionary of their equivalent synonyms and antonyms whereas corpus-based method deals with a seed list of opinion words, and then finds other opinion words in a large corpus to get opinion words with context specific orientations.[10]

#### VII. Results

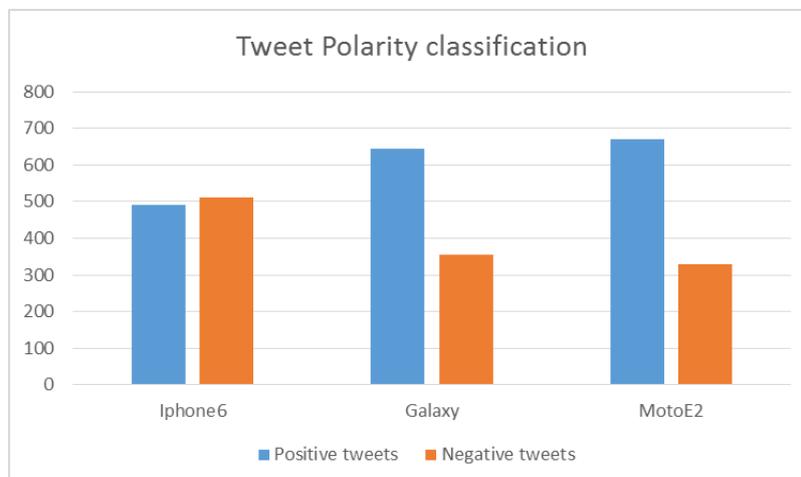
There is no large public available data set of Twitter tweets with sentiment, so we use Twitter API to collect data. The Twitter API has a parameter that specifies in which language you want to retrieve tweets and we set this parameter to English. We acquire 3000 tweets of three distinct brands each i.e. iPhone6, Samsung galaxy, motoE2. Then we performed preprocessing on tweets.

Tweets of these different mobile phones are downloaded using twitter API. Preprocessing of these tweets is done based on removing punctuations, white spaces, converting to lower case. Comparing with positive and

negative word dictionary. Then polarity of these tweets based on positive, negative and neutral values is calculated as mentioned in Table1 and chart1.

Sr.No.	Name of Dataset	Tweets Before Preprocessing	Positive Tweets	Negative Tweets
1	Iphone6	1000	490	510
2	Galaxy	1000	645	355
3	MotoE2	1000	670	330

**Table 1:** polarity classification of tweets



**Chart1:** polarity classification of tweets

### VIII. Conclusion and future work:

Twitter has become an important resource of customer reviews which helps to understand their opinions, comments about various products, companies and their services. These tweets are analyzed to find their polarity as positive, negative or neutral. By applying machine learning algorithm we get classification of these tweets and feature extraction. It helps business organizations to develop good policies for customers by using decision support system with recommendations about products. The challenges in sentiment analysis area currently there is a need to develop algorithms and/or systems which are domain independent. In future we our focus is to build a model which will help in analyzing customer opinion and predict future plans for business organization for quality improvement.

### References

- [1]. A PappuRajan , S.P.Victor, Web Sentiment Analysis for Scoring Positive or Negative Words using Tweeter Data International Journal of Computer Applications (0975 – 8887) Volume 96– No.6, June 2014
- [2]. Preety, Sunny , Dahiya, “SENTIMENT ANALYSIS USING SVM AND NAÏVE BAYES” International Journal of Computer Science and Mobile Computing, Vol.4 Issue.9, September- 2015, pg. 212-219
- [3]. Rahul Rajput, Arun Kumar Solanki , “Review of Sentimental Analysis Methods using Lexicon Based Approach” , International Journal of Computer Science and Mobile Computing, Vol.5 Issue.2, February- 2016, pg. 159-166
- [4]. Amit G. Shirbhate, Sachin N. Deshmukh , “Feature Extraction for Sentiment Classification on Twitter Data”International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064
- [5]. Jonathon Read, “Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification”, In ACL, The Association for Computer Linguistics 2005,pg.43-48.
- [6]. Alec Go, Lei Huang, RichaBhayani, “Twitter Sentiment Analysis”, Final Projects from CS224N for Spring 2008/2009 at the Stanford Natural Language Processing Group.
- [7]. Alexander Pak, Patrick Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining”, pg.1320-1326.
- [8]. Luciano Barbosa, JunlanFeng, “Robust Sentiment Detection on Twitter from Biased and Noisy Data”, Proceedings of the 23rd International Conference on Computational Linguistics: Posters 2010, pg.36–44.
- [9]. ApoorvAgarwal, BoyiXie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, “Sentiment Analysis of Twitter Data”, Proceedings of the workshop on Languages in Social Media, LSM’11, pg.30-38.
- [10]. Shailesh Kumar Yadav, “Sentiment Analysis and Classification: A Surve”, International Journal of Advance Research in Computer Science and Management Studies Volume 3, Issue 3, March 2015, pg.113-121
- [11]. Aizerman M, Braverman E, Rozonoer L, “Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning”, 1964, pg.821–837.
- [12]. SnehaVishwakarma, Suresh Kumar, “A Lexical Approach for Tweets Sentiment Classification”, International Journal of Scientific Research volume: 4, Issue: 10, ISSN No. 2277-8179, pg.242-245.