

# Sentiment Analysis of Twitter Data

**Priyanka Dongre**

*DattaMeghe College of Engineering, Airoli, Navi Mumbai*

**Dr. S. D. Sawarkar**

*Principal, DattaMeghe College of Engineering, Airoli, Navi Mumbai*

## **ABSTRACT**

*Sentiment analysis can provide measurable insights through summarization, keyword analysis and clustering. We measure sentiment using a binary choice keyword algorithm. A multi-knowledge based approach is proposed using, Self-Organizing Maps along with sentiment polarity in order to model sentiment. We develop a visual model to express a sentiment concept vocabulary and then apply this model to maximums and minimums in the time series sentiment data. The results show actionable knowledge can be extracted in real time. Twitter Data Twitter has developed its own language conventions. The following are examples of Twitter conventions.*

## **KEYWORDS**

*(Twitter data, Tokenization, Lexicons, Text Mining, Sentiments, Hive, Big data, Filtering)*

## **INTRODUCTION**

As the usage of social media has been increased with a high rate, the data which is shared on social media by the users is analyzed to get the information about the people. Many applications of sentiment analysis have been implemented to know the different sentiments of the users. People share their status on social media which has information about their movie review by posting expression in the form of negative, positive feelings. By extracting such user opinions and to carry out the analysis on the recorded results, which can give the movie review rate for a given period of time.

## **LITERATURE SURVEY**

There are many research work published in the area of sentiment analysis and classification, With this approach, we are able to train classifiers with Big Data within a few minutes. Thank to this we have evaluated in total 720 different configurations (each trained on the big data described in III). This work has also increased dimensional space using n-grams and their optimization was used.[6] A text may change its semantic subject to different contexts. For example “that scene made me cry” post may fall under two different moods i.e. happy or sad. The person cried of happiness after watching the success of his or her ward or cried of sadness after watching an accident. So it really depends on the context.[7]

## **PROPOSED METHODOLOGY**

The phases involved in this methodology are discussed below[3]:

**a) Twitter Authentication:** Before mining any data from Twitter using APIs, we have to authenticate with twitter using an application created on twitter. Once the application is

created, we get access to consumer key, consumer secret, access token, access secret using which the API has to authenticate itself with Twitter Authentication server.

**b) Twitter Data Mining(hashtag extraction):** Once API is authenticated with Twitter Authentication service, a token is generated and is made available to API for every transaction with twitter server. Using this token, tweets are mined using hashtags. We use searchTwitter() function to mine the data.

**c) Sentiment Analysis and Scoring the tweets:** Here the mined Tweets are obtained from .csv file and the sentiment analysis function performs the following task:

- Considers alphanumeric characters
- Removes punctuation
- Removes control characters
- Removes digits
- Converts the case to lower
- Splits the tweet sentences into words
- Compares words to the dictionaries of positive negative terms
- Scores the tweets

Tweets are rated based on difference between no. of positive words and no. of negative words.

- i) If Score < 0, this means that the sentence has an overall 'positive opinion'.
- ii) If Score > 0, this means that the sentence has an overall 'negative opinion'.
- iii) If Score = 0, then the sentence is considered to be a 'neutral opinion'.

Score = Positive score - Negative score

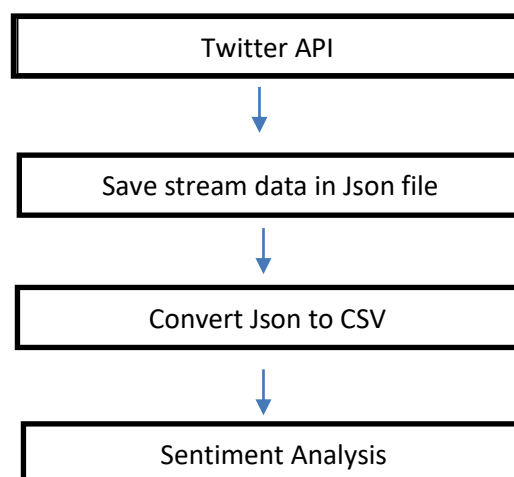
Following code snippet is used to analyse the sentiment of mined tweets:

```
score.sentiment = function(sentences, pos.words, neg.words, progr)
```

**d) Text Mining:** Once the sentiments of the tweets are analyzed, text mining is used to generate Word- cloud is based on various filter conditions defined by the user. Word cloud can be generated by using a function word cloud defined in wordcloud package.

We have used Hive (big data terminology) for faster retrieval of tweets with great accuracy, as following figure shows that.

#### PROCESS FLOW OF TWITTER DATA



**Fig 1:Process Flow**

We are doing sentiment analysis of twitter data of movies using Hive, reason behind choosing hive is it supports SQL based queries, programmability, interoperability with other database tools. In this case, we are analyzing the sentiments of movie through Twitter data so that it will be easy to identify review of movie. For this purpose hive is used with lexicon based approach with the help of naïve bayes algorithm for accuracy.[10]

As per mentioned in fig.1 get the credentials of twitter API i.e consumers key, consumers secret key, access token, access secret token. All the twitter data collected for movie will be saved in Json file. To collect the streaming data we will use tweepy. The necessary information from each tweet will be saved in csv file. From the tweets text the sentiment analysis will be done using textblob.

**EXPERIMENTAL RESULTS**

Twitter Sentiment Corpus -Collection of 5513 tweets

collected for 4 Movies : Bahubali, Ravan

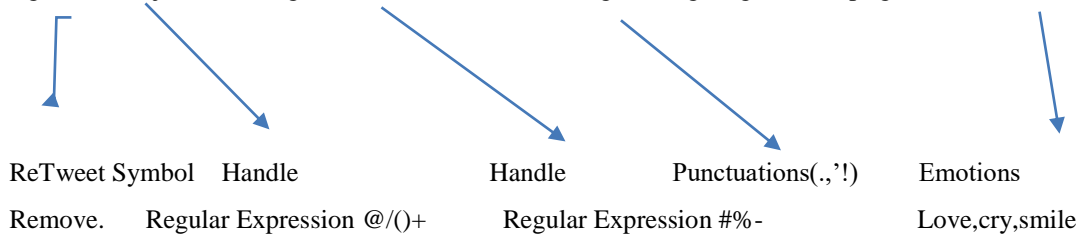
- o Each tweet is classified as Positive or Negative

**Table 1. Twitter Sentiment Corpus**

Class	Count	Example
Negative	2214	#Bahubali is extra dramatic.
Positive	2168	#Singham is best film.

**Preprocessing : Reducing the size of Dataset**

e.g. RT@ssrajamouli..congrats..#bahubali is outstanding.. Will give goosebumps guaranteed.. <3



**Table 2.Results of reduction in feature set and stemming**

Twitter Sentiment		Standard Corpus		Both		
Preprocessing	Words	%	Words	%	Words	%
None	19128		15910		31832	
Hashtags	18649	<b>97.50</b>	15550	<b>97.74</b>	31223	<b>98.09</b>
Handles	17118	<b>89.49</b>	13245	<b>83.25</b>	27383	<b>86.02</b>
URLs	16723	<b>87.43</b>	15335	<b>96.39</b>	29083	<b>91.36</b>
Emotions	18631	<b>97.40</b>	15541	<b>97.68</b>	31197	<b>98.01</b>
Punctuations	13724	<b>71.75</b>	11225	<b>70.55</b>	22095	<b>69.41</b>
Repetitions	18540	<b>96.93</b>	15276	<b>96.02</b>	30818	<b>96.81</b>
All	11108	<b>58.07</b>	8646	<b>54.34</b>	16981	<b>53.35</b>

## CONCLUSION AND FUTURE WORK

Sentiments are the words or sentences that represent view or opinion that is held or expressed that can be positive, negative. We proposed a hybrid approach involving both corpus-based and dictionary-based techniques, which will find the semantic orientation of the sentiments words in tweets. We will also consider features like emoticons, neutralization, negation handling and capitalization as they have recently become a huge part of the internet language. The accuracy of the model is highly dependent on the way pre-processing is done.

Future expertise can improve sentiment analysis method. Exploring further methods of quantitative analysis such as summarization.

## REFERENCES

1. Yuji Yano, Tomonori Hashiyama, "Behavior Extraction from Tweets using Character N-gram Models" 2014 IEEE
2. Wei Yen Chong, Bhawani Selvaretnam "Natural Language Processing for Sentiment Analysis" 2014 4th International Conference on Artificial Intelligence
3. Dr Nirmala C R, Roopa G M, "Twitter Data Analysis for Unemployment Crisis"
4. Shoiab Ahmed, AjitDanti, "A Novel Approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using Web Data" ©2015 IEEE
5. Pankaj Gupta, Ritu Tiwari and Nirmal Robert , " Sentiment Analysis and Text Summarization of Online Reviews: A Survey" @2016 IEEE
6. Lukas Povoda, RadimBurget, "Sentiment Analysis Based on Support Vector Machine and Big Data" ©2016 IEEE
7. Govin Gaikwad, Prof. Deepali J. Joshi, "Multiclass Mood Classification on Twitter Using Lexicon Dictionary and Machine Learning Algorithms"
8. UmmeAymunSiddiqua, Tanveer Ahsan. "Combining a Rule-based Classifier with Weakly Supervised Learning for Twitter Sentiment Analysis" @2016 IEEE
9. MondherBouazizi, "A Pattern-Based Approach for Sentiment Analysis in Twitter"
10. William B. Claster, "Modeling Sentiment from Twitter Tweets using Naïve Bayes"