*International Journal of Engineering & Technology,* 7 (2.31) (2018) 80-84

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# Sentimental analysis on social media data using R programming

**Mandava Geetha Bhargava[1]*, Duvvada Rajeswara Rao[2]**

*[1]M.Tech Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Educational Foundation (KLEF), Andhra Pradesh, India.*
*[2]Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Educational Foundation (KLEF), Andhra Pradesh, India.*

**Abstract**

Sentimental Analysis is an ongoing research field in Text Mining Arena to determine the situation of market on particular entity such as Product, Services...Etc. and it can be called as computational treatment of reviews, subjectivity and sentiment of text. Cryptocurrency can be explained as a type of digital estate and devised to mechanize as a form of trade and exchanges that uses cryptography as an encryption technique to secure the transactions and acts as decentralized controlled transaction which is opposed to centralized transactions. Cryptocurrency are a type of virtual currency, digital currency and alternative currency, On basis of categorical, there are different architecture and security protocols which are used in the cryptocurrencies to secure transactions, the different types of cryptocurrency are available in the market such as Bitcoin, Litecoin, and Namecoin…etc. This paper focuses on survey on different types of sentimental analysis methods and main contribution of this paper include sentimental analysis of social media data on different types of cryptocurrencies on basis of categorical and different terms of cryptocurrency such as Cryptocurrency, virtual currency, digital currency and discussed on trends of crypto currency in present market.

***Keywords:*** *Text Mining, Polarity, Emotions, Bitcoin , Litecoin*

## 1. Introduction

Sentimental Analysis is an computationally categorizing and identifying opinions expressed in a piece of text which was to determine the user's, author's, and writer's reviews towards a particular product, topics...etc respectively and it can be also called as Text Mining. The meaning of Text Analysis and sentimental analysis is interchangeable and both words express mutual meaning, however some researchers told that Opinion Mining and Sentimental Analysis have negligible different views. Opinion Mining analyses and extracts the opinions of people on particular entity such as product, events...Etc. Sentimental Analysis identifies the sentiment in form of text which was given by user's or author's then analyses it, hence the target of the sentimental analysis is to identify sentiments which was expressed and classify them into different rate of polarity such as a rating system. Cryptocurrency is a digital estate designed work as a channel of exchange that uses cryptography to secure its transactions, to verify the transfer of estate and to control the creation of additional units .these are subsets of digital currencies and are also subset of virtual currencies and alternative currencies, There are different types of cryptocurrency as follows: Bitcoins, Litecoins, Namecoins,Swiftcoins,Bytecoins,Peercoins, and etc. There are cryptocurrency exchanges which as a channel for trading other assets such as different digital currencies and conventional flat money, The different types of cryptocurrency exchanges are Btc.sx, CEX.IO, BTCChina, Coinbase, Luno...etc. In above informed cryptocurrency, we focused on BitCoins, LiteCoins, and PeerCoinsfor text mining.In this paper we focused on the present trends of cryptocurrency by means of Text Mining and comparative study of different algorithms as shown in figure 1 and follows: 1) capturing of data,2) Data Processing ,3) Sentiment Classification, 4) Sentiment Polarity 5) Visualization by different Techniques

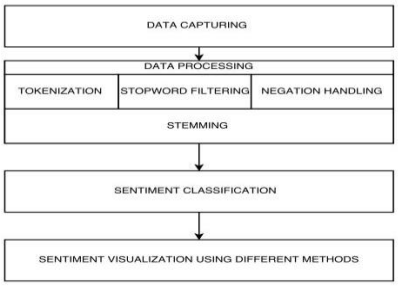


**Fig. 1:** Process of sentimental analysis

withhelp of application interfaces (API) ,in this paper we used the social media data of Facebook and twitter datausing API's and procedure to capture the data from social media site is explained in methodology section .Preprocessing is an method to clean and sets in order the captured data and then the different captured data are combined. Data Processing: It is an important step for sentimental analysis of any entity where the captured data contains hashtags, emoticons and other punctuations are filtered. Sentiment Classification is astep for classifying the tweets or posts from social media site such as twitter and facebook respectively with help of different algorithms, where it helps to classify sentimentsinto differentpolarity's and emotion's. Sentiment Polarity is a step for assigning different polarity as per each sentiment with help of used algorithm such as positiveness and negativeness.Sentiment Visualization is animportant step for Visualization of Polarity and Emotions which was classified by selected algorithm and where the end-user or user can understand

the phenomena of selected entity in market arena. In this Paper, we are going to discuss about Sentimental Analysis procedure for Present trends of Cryptocurrency and Comparative study of different algorithms such as Naïve Bayes, SVM as shown in Figure 2.

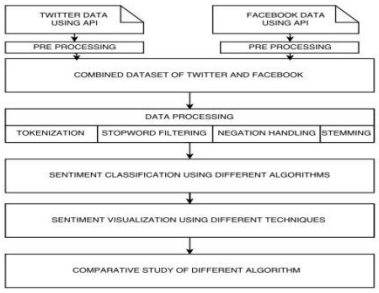


**Fig. 2:** Sentimental analysis process on trends of cryptocurrency

There are Different types of Classification methods for Sentimental analysis are as follows: 1) Machine Learning Based method, 2) Lexicon Based method, 3) Hybrid Based method and the features of sentiment classification methods are shown in table.1

## 2. Research background

**Devika M D et al [1]** studied on comparative study of different approaches in sentiment analysis by means of comparing the different approaches and methods of sentimental analysis such as machine learning approach, N-gram Sentimental Analysis ,Ruled and Lexical approaches, Maximum Entropy Classifier,support vector machine and Naives Bayes Method…etc. and they concluded that the machine learning methods such as Support vector machine, Maximum Entropy and Naives Bayes Methods are more efficient and semantic analysis of text makes process easier while doing sentimental analysis of social media.

**Ravendra Ratan Singh Jandail et al [2]**studied on survey of sentimental analysis and opinion mining with respect to needs and requirements of organization and customer respectively, compared the techniques of data mining and machine learning in sentimental analysis and concluded that opinion mining and sentiment analysis is an emerging research field of data mining and natural language processing

**Table 1:** Features of Different Classification Methods

| SENTIMENT CLASSIFICATION Methods | | FEATURES |
|---|---|---|
| MACHINE LEARNING | Bayesian Networks<br>Naive Bayes Classification<br>Maximum Entropy<br>Neural Networks<br>Support Vector Machine | Term Presence and Frequency<br>Part of Speech Information<br>Negations<br>Opinion Words and Phrases |
| LEXICON BASED | Dictionary Based Techniques<br>Novel Machine Learning Techniques<br>Corpus Based Techniques<br>Ensemble Techniques | Manual Construction<br>Corpus based<br>Dictionary Based |
| HYBRID | Machine Learning<br>Lexicon Based | Sentiment Lexicon constructed using public resources for initial sentiment detection<br>Sentiment words as features in machine learning method |

**Narahari P Rao et al [3]** studied on real time opinion mining application of twitter data using integration workflow of twitter API, Flume, HDFS (Hadoop Distributed File System), Hive and Oozie.
concluded that their architecture can useful for identifying spam accounts and clusters of keywords, their architecture can supports other social media such as Facebook, personal blogs, movie reviews...etc.

**Tajinder Singh and Madhu Kumari [4]** done a work on text pre-processing in twitter sentiment analysis with help of text normalization and n-gram to find and filter the different slangs and tags respectively, their used Support Vector Machine (SVM) for evaluation and measurement of the impact of their proposed scheme on the sentimental classification task as classifier and concluded that their proposed scheme allows to classify sentimental data from noises easier.

**Alexander Fillbrunn et al [5]**studied on reproducible cross domain analysis of life science data using knime tool about different extensions and work flow tools for simplifying data exploration, analysis and visualization of life science data and concluded that knime makes integration of different workflows and extensions easier to data exploration, analyse, visualize the life science data.

**Banan Awrahman and Bilal Alatas[6]**Studied on opinion mining and sentimental analysis of social networks using knime tool and concluded that there are various aspects and types of techniques to handle different types of text format in social network data.

**Neeti Sangwan et al [7]** studied on text analytics using R language by means of twitter as a platform for text mining and CART and Random forest are used for analysing the data and monitoring and removing the spams, concluded that text analytics is the field pertaining to gather data to analyse and study the general public opinion.

**Pooja Khanna et al [8]**studied on sentimental analysis with Twitter Data using Rby meansof capturing the social media data from database with help of Application Interface .Analysed and Visualized the data by means of R language with help of graphs and concluded that opinion mining makes the public attitudes to generalise the prevailing trends of the market.

**Praniti R.Thanvi et al [9]**studied on opinion mining of political reviews using adjective-adverb-verb-noun ( AAVN) combinations and Naïve Bayes technique ,concluded that their application makes users easier to understand the opinions of relevant political sites and efficient to make opinion mining.

**K.Arun et al [10]**studied on sentiment analysis on demonetization tweets in twitter social media in India using R language and made a analysis and visualization on several tweets such as Demonetisation, Digital Payments, Operation Clean Money, Income Tax Payments and made word cloud and different types of plots by means of filtering the captured data from several noises and concluded that most of people accepting is demonetization as positive.

## 3. Methodology

In previous section, we discussed about different methods to analyze the sentiments. In this section we are going analyze the real time data by capturing the tweets and posts from social media database such as Twitter and Facebook as shown in below Algorithm.1. Firstly, we have to Create Developer Account in Facebook and Twitter for Creating Applications to capture real time data and then create an application using developer website of twitter, similarly create an application from developer website of Facebook. Generate the Application Interface's or Consumer's Keys and Access key for capturing required content from social media. After captured of data from social media database, it is processed in different steps such as Tokenization, Removal of Stopwords, Negation Handling, and Stemming. Tokenization is a process of formatting the part of speech into individual words,

Phrases, keywords…etc. Stopwords are basically set of common used words such as a, an,the,to, of etc. which we want to exclude while analyzing text. Negation Handling is a process of handling Negative polarity sentence or document from positive and neutral sentence or document. Stemming is an technique to stem different words into original word, so that the words would look normal and sentiment or corpus are classified by Naives Bayes and Supported Vector Machine Algorithm, Emotions are assigned to different documents by means of word dictionary and Polarity is assigned to each and every sentences or document for classifying as positive , neutral and negative where it makes easier to visualization of analyzed data by means of classification by emotions , word clouds and Classification by Polarity. Comparative Study is made on Naïve biased and Support Vector Machine algorithms by means of efficiency and confusion matrix as shown in next section.

Algorithm 1:

Step 1: Create $P_{api}$

Step 2: Generate $K_C$, $K_A$

Step 3: Start

Step 4: Capture the $D_i$ and $D_{ii}$using $K_C$ and $K_A$ Respectively of selected micro blogger site

Step 5: Pre-Process the captured $D_i$ and $D_{ii}$

Step 6: Combine $D_i$ and $D_{ii}$as $D_{iii}$

Step 7: Process $D_{iii}$ using different techniques

Step 8: Classify $D_{iii}$ with help of different algorithms

Step 9: Visualize the $D_{iii}$ using different techniques

Step 10:Compare Efficiency of Used Algorithms

As shown in Algorithm, First we created the application Interface $P_{api}$ for capturing tweets and posts from microblogging (social media) sites and Access Keys ($K_A$) and Consumer Keys ($K_C$) of Applications are generated with respective of micro-blogging sites i.e. twitter and facebook, and by means of authorization and authentication, we are capturing the data $D_i$ and $D_{ii}$ from both micro-blogging sites. The Captured Data are preprocessed by means of stemming and storing of text data and there are combined as $D_{iii}$. There are processed using different techniques if required and the processed data is classified by means of different algorithms such as Naïve Based and Support Vector Machine and there are visualized in terms of emotions, polarities and confused matrix of used algorithms. Comparison and explanation of both algorithms is done in next section.

## 4. Results and discussion

After Implementing and processing of the captured data from social media as shown in above section, we got following results as shown below. The results deals about the Sentimental Analysis of terms such as cryptocurrency, digital currency, bitcoins, litecoins and virtual currency as discussed from previous section, Fig.3. Shows about the rate of polarity with respect to Source of data and type of algorithm used on term Cryptocurrency, Fig.4. Shows about the rate of emotions with respect to source of data and type of algorithm used on term Cryptocurrency, similarly Fig.5, Fig.6, Fig.7,Fig.8, Fig.9,Fig.10,Fig.11,Fig.12 shows about the rate of polarity and emotions with respect to source of data and type of algorithms.

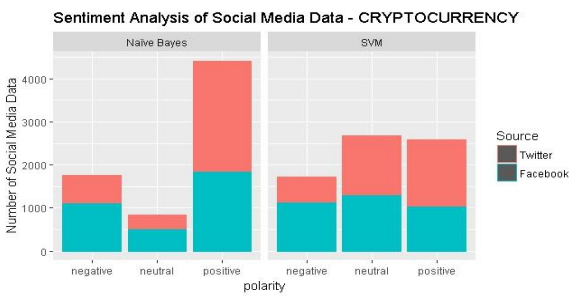


**Fig. 3:** Sentiment analysis of social media data about cryptocurrency (classification by polarity)

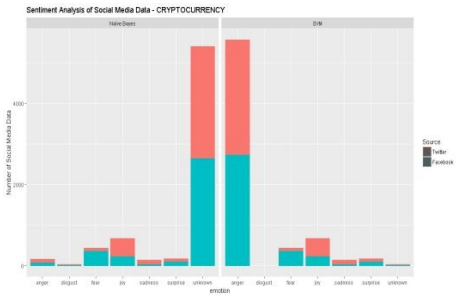


**Fig. 4:** Sentiment analysis of social media data about cryptocurrency (classification by emotion)

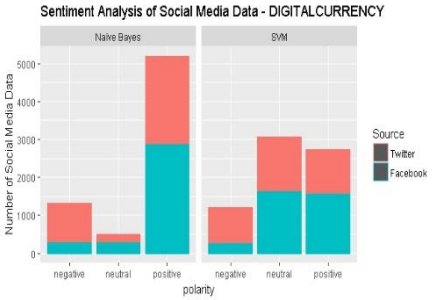


**Fig. 5:** Sentiment analysis of social media data about digital currency (classification by polarity)

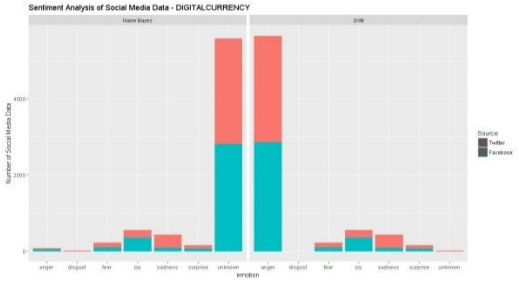


**Fig. 6:** Sentiment analysis of social media data about digital currency (classification by emotion)

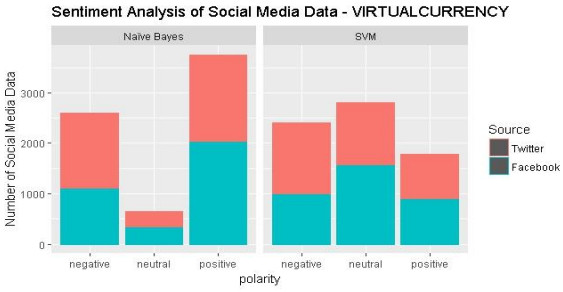


**Fig. 7:** Sentiment analysis of social media data about virtual currency (classification by polarity)

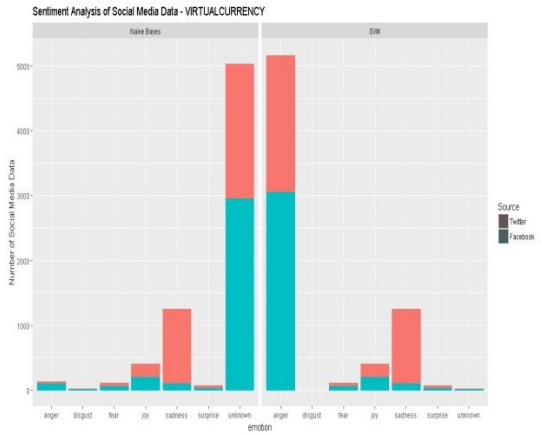


**Fig. 8:** Sentiment analysis of social media data about virtual currency (classification by emotion)

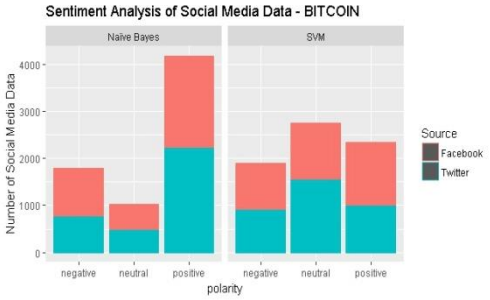


**Fig. 9:** Sentiment analysis of social media data about bitcoin (classification by polarity)

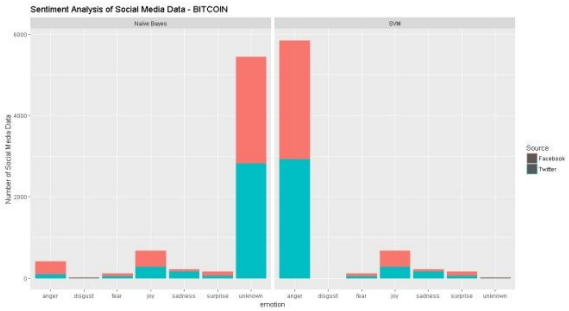


**Fig. 10:** Sentiment analysis of social media data about Bitcoin (classification by emotion)

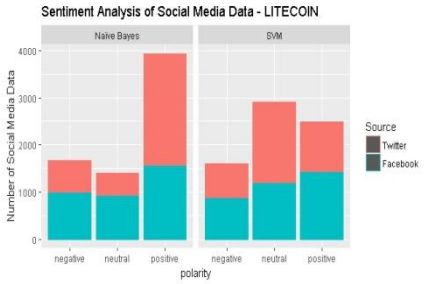


**Fig. 11:** Sentiment analysis of social media data about litecoin (classification by polarity)

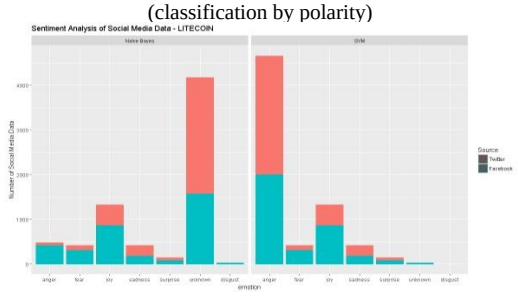


**Fig. 12:** Sentiment analysis of social media data about litecoin (classification by emotion)

As discussed about used algorithms in methodology section such as Support Vector Machine and Naïve Bayes are compared as shown in below tablewith respect to different topics used in sentimental analysis and accuracy of used algorithm in sentimental analysis as shown in below table.2

**Table 2:** Accuracy of Algorithms Used in Sentimental Analysis with Respect to Dataset

| Topic Name Related to Dataset | Accuracy of Algorithms | |
|---|---|---|
| | Naïve Bayes | Support Vector Machine |
| Bitcoin | 59.3% | 35.8% |
| Litecoin | 51.9% | 38.2% |
| Cryptocurrency | 57.5% | 35.9% |
| Virtual Currency | 48.3% | 36.4% |
| Digital Currency | 69.6% | 39.9% |

## 5. Conclusion and further work

This paper discussed about sentimental analysis on cryptocurrency and its various product in present market arenaand discussedon various methods and features on sentimental analysis of social media data using application interfaces, from previous section we conclude that the term called ‘cryptocurrency’, ’digital currency’ has more positive in social media with respect to both algorithms and the term called ‘virtual currency’, ’bitcoin’ and ‘litecoin’ has little bit neutrality with positivewith respect to both algorithms and Naïve Bayes has more Accuracy than Support Vector machines in sentimental analysis of social media data,we can make real time applications with sentimental analysis on social media if the constraints with application interfaces are resolved.

## References


[1] Devika MD, Sunitha C & Ganesh A, “Sentimental Analysis: A comparative Study on Different Approaches”, *procedia computer Science*, Vol.87, (2016), pp.44-49.

[2] Jandail RRS, Sharma P & Agrawal C, “A Survey on Sentiment Analysis and Opinion Mining: A need for an Organization and Requirement of a Customer”, *IJETAE*, Vol.4, No.3,(2014), pp.17-24.

[3] Rao NP, Nitin Srinivas S & Prashanth CM, “Real Time Opinion Mining of Twitter Data”, *International Journal of Computer Science and Information Technologies(IJCSIT)*,Vol.6, No.3, (2015).

[4] Singh T & Kumari M, “Role of Text Pre-Processing in Twitter Sentiment Analysis”, *Procedia Computer Science*, Vol.89,(2016), 549-554.

[5] Fillbrunn A, Dietz C, Pfeuffer J, Rahn R, Landrum GA & Berthold MR, “KNIME for reproducible cross-domain analysis of life science data”, *Journal of biotechnology*, Vol.261, (2017), pp.149-156.

[6] Awrahman B & Alatas B, “Sentimental Analysis and Opinion Mining within the social networks using Konstanz Information Miner”, *Journal of Telecommunication, Electronic and Computer Engineering* ,Vol.9, (2016), pp.15-22.

[7] Sangwan N, Malik V, Sunesh & Sangwan S, “A Survey for Text Analysis Using R language”, *Advances in Computational Sciences and Technology*, Vol.10, (2017), pp.2173-2182.

[8] Khanna P, Kumar S, Mishra S & Sinha A, “Sentiment Analysis: An Approach to Opinion Mining from Twitter Data Using R”, *International Journal of Advanced Research in Computer Science*, Vol.8, (2017), pp.252-256.

[9] Thanvi PR, Sontakke NS, Waghmare SR, Patel ZS & Gavhane S, “Sentimental Analysis for Political Reviews using AAVN Combinations”, *International Research Journal of Engineering and Technology (IRJET)*,Vol.04, (2017), pp.72-74.

[10] Arun K, Srinagesh A & Ramesh M, “Twitter Sentiment Analysis on Demonetization tweets in India Using R language”, *International Journal of Computer Engineering In Research Trends*,Vol.4, No.6, (2017), pp.252-258.

[11] sandhyarani K & kavitha D, “Sentiment Analysis of Social Media Data”, *International Journal of Emerging Trends and Technology in Computer Science(IJETTCS)*,Vol.6, No.5,(2017), pp.151-156.

[12] Hutto CJ & Eric Gilbert, “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text ”, *Proceedings of the Eighth International AAAI Conference on weblogs and social media*, (2014), pp.216-225.

[13] Abdul-Mageed, M Diab M & Kubler S, “SAMAR: Subjectivity and Sentimental Analysis for Arabic Social Media”, *Computer Speech and Language*, Vol.28, No.1, (2014), pp.20-37.

[14] Nguyen TH, Shirai K & Velcin J, “Sentimental analysis on Social Media for stock movement prediction”, *Expert Systems with Applications*, Vol.42, No.24, (2015), pp.9603-9611.

[15] Sinha H & Kaur A, “A detailed survey and comparative study of sentiment analysis algorithms”, *2nd International Conference on Communication Control and Intelligent Systems (CCIS)*,(2016), pp. 94-98.

[16] Wang H, Hanafy A, Bahgat M, Noeman S, Emam OS & Bommireddipalli VR, “A System for Extracting Sentiment from Large-Scale Arabic Social Data”, *First International Conference on Arabic Computational Linguistics (ACLing)*, (2015), pp.71-77.

[17] Kikuchi S & Bhalla S, "Data model for reputational analysis for combined medias social networks and IPTV", *IEEE 3rd Global Conference on Consumer Electronics (GCCE)*, (2014), pp. 51-52.
[18] Kumar S, Singh P & Rani S, "Sentimental analysis of social media using R language and Hadoop: Rhadoop", *5th International Conference on Reliability Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, (2016), pp.207-213.
[19] Yang N, Cao S & Zhang S, "Data analysis system for online short video comments", *IEEE/ACIS 15th International Conference on Computer and Information Science*, (2016), pp.1-6.
[20] Parveen H & Pandey S, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm", *2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, (2016), pp.416-419.
[21] Soni AK, "Multi-lingual sentiment analysis of Twitter data by using classification algorithms", *Second International Conference on Electrical Computer and Communication Technologies (ICECCT)*, (2017), pp.1-5.
[22] Zhang L, Liu B & Vaithyanathan S, "Identifying noun product features that imply opinions", *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers*, Vol.2, (2011), pp.575-580.