International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 14 (2018) pp. 11525-11531

Sentiment Analysis and Prediction of Online Reviews with Empty Ratings

Sasikala P*[1], L.Mary Immaculate Sheela[#2]

**Research Scholar, Department of Computer science,*
Mother Teresa Women's University, Kodaikanal, India.

[#]Pentecost University College, Accra, Ghana.

Abstract

Online reviews are a category of product information created by the users based on personal handling experience. Online shopping websites endow with platforms for consumers to review products and carve up opinions. Sentiment analysis or opinion mining is nothing but classification of emotions in the reviews text into positive, negative and neutral. Opinion mining is a method of information extraction from text processing to improve or develop the business work by review analysis. The problem is most of the comments from customer reviews about the products are contradicted to their ratings. Many customer will post their comments and forgot to rate the product or not engrossed to rate it. In this work we have designed a classifier model which accepts all the reviews and group them into two categories as reviews with ratings and reviews with blank or empty ratings. Further prediction of sentiments using various classifiers is done for the reviews without ratings.

Keywords: Classifier, Online Reviews, Sentiment Analysis, Wordcloud.

INTRODUCTION

Many consumers rely on online reviews for direct information to make purchase decisions. However, a large number of reviews for just one single product have made it impractical for consumers to read all the reviews and assess the true quality of a product. In addition, the quality and the helpfulness of each review also fluctuate. The large quantity of the reviews and their serrated quality make it tough for consumers to differentiate between useful and useless reviews. Public opinion plays a vital role in business organization to market the products, venture new opportunities and for sales prediction. A large amount of data can be analysed and prediction of opinion is possible using sentiment analysis techniques which helps the customers and business organisation.

The unrefined online data comprise of multiple flaws which obstruct the sentiment analysis process. The flaws include unguaranteed opinions, irrelevant opinions to the topic, spam and fake opinions. Ground truth of such online data is not always available [1].The ground truth is tagging the data with sentiments as positive or negative. Amazon is one of the biggest online vendors in the world. Data used in this paper is online consumer reviews of products collected from Amazon [2] in the time period of February 2012 to July 2017. The flaws prevailed over by verification before posting the reviews and ground truth is ratings mentioned in the scale of 1 to 5.

This paper deals with basic problem of prediction of ground truth. Most of the customers had written the reviews and left the ratings as empty. The reviews with the comments like "worst" and "average", but the ratings were rated in the scale of above 4.0 which is contradict to their comments. In this paper we have predicted the sentiments for the reviews with empty or blank ratings using various sentiment analysis classifiers. The experimental results of the classification model compared and classified the review text into positive and negative. We have simulated a model implemented in Python which accepts the reviews texts and categorise them into positive and negative with probabilistic value.

RELATED WORK

Xing et al. [1] had proposed a work on product reviews collected from amazon to identify the negation phrases. Sentence level and review level classification of data is performed for the data collected from February to April 2014. Aashutosh Bhatt et al. [3] used reviews of iPhone 5 extracted from Amazon website and suggested a rule based extraction of product feature sentiment analysis. POS technique is implemented to each and every sentence level and the results are shown in charts. Ahmad Kamal [4] used supervised and rule based techniques to mine the opinions from online product reviews. Bhumika et al.[5] compared the machine learning models and shown the comparison of efficiencies of these models for Twitter data .

PROPOSED WORK

Sentiment classification is to select and extract the text features. Feature selection in sentiment analysis is collecting the information from reviews in web and performing the following steps.

Data Preparation: The data preparation step will pre-process the data and removes all the non-textual information and tags. Data pre-processing performs cleaning of data by removing

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 14 (2018) pp. 11525-11531

the information like review date and name of the reviewer which is not required for sentiment analysis.

Review Analysis: Bag of words model is used to categorize documents and frequency of occurrence of words is extracted for training the classifier.

S**entiment Classification**: Classifies the extracted words as positive or negative.

The architectural view of the system is given in fig.1.

In machine learning, classification is used to classify the given content into a precise set based on a training set of data containing observations whose category is known in advance. Classifier algorithms can be used to categorize sentiments of review based on words. The specific words in the language are categorized in advance for their positive or negative sentiments. Classification is an instance of supervised learning. Training set has correctly identified observations. Classifier algorithms are used to create clusters from the uncategorized unsupervised data based on similarity from the training data set.

The data is collected from amazon customer reviews list of over 34,000 consumer reviews for Amazon products like the Kindle, Fire TV Stick, DVD and Electronic items from Feb 2012 to July 2017.The dataset includes basic product information, rating, review text, reviews username and reviews title. The top 5 reviews with ratings are shown in (Table 1). The rows with empty or null ratings are represented in table2.

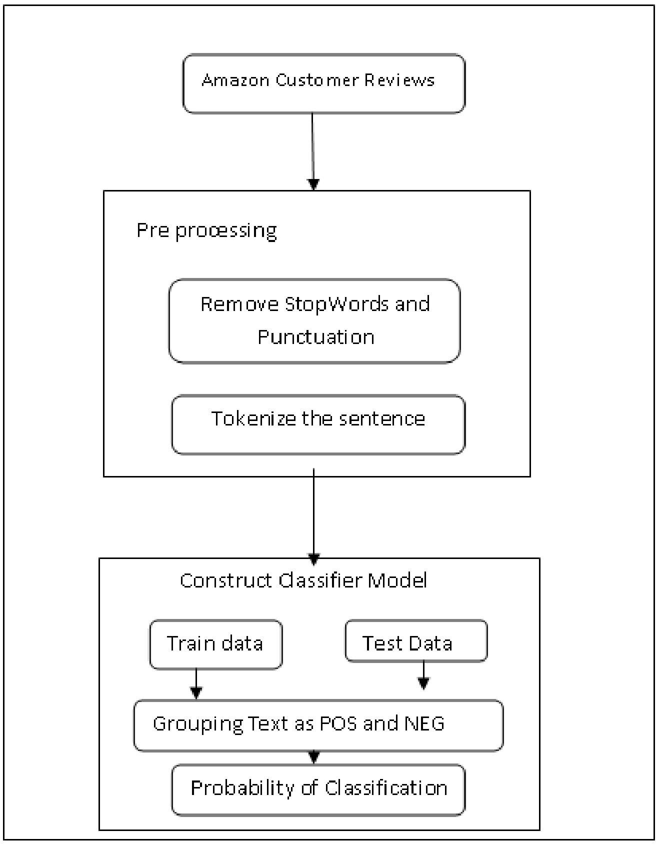

Figure 1. Architecture of the proposed system.

Table 1. Reviews with ratings

	reviews. Rating	**reviews.text**	**reviews.title**	**reviews.username**
0	5.0	This product so far has not disappointed. My c...	Kindle	Adapter
1	5.0	great for beginner or experienced person. Boug...	very fast	truman
2	5.0	Inexpensive tablet for him to use and learn on...	Beginner tablet for our 9 year old son.	DaveZ
3	4.0	I've had my Fire HD 8 two weeks now and I love...	Good!!!	Shacks
4	5.0	I bought this for my grand daughter when she c...	Fantastic Tablet for kids	explore42

Table 2. Reviews with null values

	reviews.rating	reviews.text	reviews.title	reviews.username
2886	NaN	The Kindle is my first e-ink reader. I own an ...	Worth the money. Not perfect, but very very go...	Jeffrey Stanley
2887	NaN	I'm a first-time Kindle owner, so I have nothi...	I Wanted a Dedicated E-Reader, and That's What...	Matthew Coenen
2888	NaN	UPDATE NOVEMBER 2011:My review is now over a y...	Kindle vs. Nook (updated)	Ron Cronovich
2889	NaN	I'm a first-time Kindle owner, so I have nothi...	I Wanted a Dedicated E-Reader, and That's What...	Matthew Coenen
2890	NaN	I woke up to a nice surprise this morning: a n...	Not the perfect do-it-all device, but very clo...	C. Tipton

Pang and Lee[6] recommended that before sentiment analysis all the objective content should be eliminated and subjectivity mining constructs valuable summaries of document sentiment. The dataset without pre-processing is used to analyse the sentiments Data is classified into positive and negative as shown in the fig 2 based on the rating .The data is not balanced and there is a vast difference between positive and negative reviews.

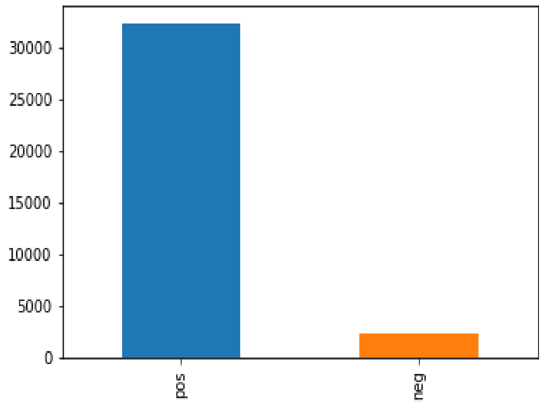

Figure 2. Classifying text as positive and negative

To balance the dataset and predict the reviews, pre-processing and feature extraction technique is used. Data is pre processed and tokenized by removing the special characters and symbols. Further the given dataset is grouped into train and test data. The classifier model is as follows:

- **Naïve Bayes**

The Bayesian Classification is a supervised statistical method for classification and contains practical learning algorithms. The posterior probability of a class can be computed using Naive Bayes model. This model works is suitable for a large data set. The use of the Bayes Theorem is to presume the chance of the inclined feature set matches to specified label. Bayes theorem provides a way of calculating the posterior probability, P(L|F), from P(L), P(F), and P(L|F). Naive Bayes classifier assumes that the effect of the value of a predictor (F) on a given class (L) is independent of the values of other predictors. The equation can be written as follows [8]:

$$P(L|F) = \frac{P(F|L)*P(L)}{P(F)} \quad (1)$$

Multinomial Naive Bayes is a focused version of Naive Bayes for text documents. Simple naive Bayes would model a document as the presence and absence of particular words, multinomial Naive Bayes explicitly models the word counts and adjusts the underlying calculations. In a multinomial event model, samples (feature vectors) represent the frequencies with which certain events have been generated by a Bayes Classifier. Bernouli model represents the occurrence or nonexistence of features in the document.

- **Logistic Regression**

Logistic Regression predicts the probability of a categorical dependent variable. The dependent has a binary variable codes as yes or no. Logistic regression works better with the large sample size. The logistic function is a sigmoid function, which takes any real nput x and outputs a value between zero and one.

$$\sigma(t) = \frac{1}{1+e^{-t}} \quad (2)$$

EXPERIMENTAL RESULTS

Naive Bayes classifier applies the probabilities of each attribute fit in to every class to make prediction. Python has natural language toolkit (NLTK) for text processing and classification. Tokenizing the text and filtering is flexible by using the toolkit. Bag of words (BOW) method is used to extract the features from the reviews. To train the classifier the words are used as features.

NaiveBayes classifier is unpredictably successful in training since its classification result may often be correct even if its probability estimates are inaccurate [7]. Data is split into test set and train set by the classifier. Train set is for training the model and test set is for executing error analysis.

The NLTK Naives Bayes accuracy and its most informative features are shown in the table 3. Randomly selected seven word's negative to positive ratio of incidences is presented in the above table. The term "warning" appears 51.3 more times as often in negative reviews as it does in positive reviews.

Table 3. NLTK Naïve Bayes Accuracy and Most Informative Features

```
NLTK Naive bayes Accuracy : 0.5533756491633006

Most Informative Features
            warning = True          neg : pos    =     51.3 : 1.0
          attempted = True          neg : pos    =     51.3 : 1.0
               logo = True          neg : pos    =     41.9 : 1.0
               nope = True          neg : pos    =     32.6 : 1.0
             savers = True          neg : pos    =     32.6 : 1.0
                opt = True          neg : pos    =     32.6 : 1.0
         repeatedly = True          neg : pos    =     32.6 : 1.0
```

The constructed model utilizes CountVector and TFIDF vector to check the test and train data. CountVectorizer converts the case, filter stop words and extracts the tokens.TFIDF (Term Frequency times inverse document frequency) increases comparatively to the number of times a word appears in the document and is equalize by the frequency of the word in the quantity. It reduces the influence of more common words like (be, at, an etc.) which occurs in all document. The accuracy of obtained by the classifiers of the model is shown in the table 4.The frequency of the features and it coefficient values are depicted in table5. Logistic regression accuracy is 93% and more compared to the other classifiers.

Table 4. Classifiers and Accuracy

Multinomial Accuracy	**: 0.928736295441**
Bernoulli Accuracy	**: 0.923542989036**
Logistic Regression Accuracy	**: 0.933496826313**

Table 5. Features and its Coefficient

	feature	coef
26451	great love	-36.6615
70852	works great year	-28.1483
53017	returning	-27.6736
61802	terrible	-26.1168
56423	slow	-26.0247
55737	shuts	-26.0008
44073	order two	-25.8462
48561	product price got	-25.6679
7559	bought gift loved	-25.4239
52972	return	-24.3509
51223	reading lightweight	-23.1062
9808	catch reading	-22.4062
51744	really shopping	-22.3478
44091	ordered several	-22.1342
71346	year old love	-21.8969
70020	wish call	-21.278
65043	usb	-21.254
14531	disappointed	-21.1124
41326	not	-20.9785
33839	limited	-20.9232
18317	every time	-20.5707
28466	holding good	-20.4839
70868	works needed	-20.4729
53005	returned item	-20.3122
41443	not buy	-20.2796
28727	honestly	-20.1628
28779	hopefully	-20.1345
33859	limited games	-19.9785
55345	setup echo	-19.9524
40836	next time	-19.8523
6	ability	17.14142
12172	convenient	17.16515
19617	far	17.27919
1259	alexa	17.42466
12601	coverage	17.67457
71023	worry	18.29401
70788	works great love	18.47615
27595	happy purchase	18.73035
41513	not disappointed	19.11561
27917	hd recommend	19.25101
5757	best	19.5087
60307	tablet great year	19.80825
5522	beat	20.21943
10609	christmas gift loved	20.41684
11912	continent	20.70625

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 14 (2018) pp. 11525-11531
© Research India Publications. http://www.ripublication.com

	feature	coef
46548	pleased	20.74292
68823	well	20.75956
24072	glad	21.54381
9137	cable	21.71459
23578	gift	21.804
2123	amazing	24.86265
4788	awesome	25.0837
18670	excellent	27.06164
36220	loves	27.22225
5246	basic amazon	27.28495
22546	fun	28.17218
45226	perfect	28.3228
15845	easy	29.88441
35297	love	36.56756
25830	great	44.09936

ROC selects a threshold for a classifier. The curve maximises the true positives and minimises the false positives. The computation used is area under curve (AUC). A particular threshold is selected and the classifiers performance at that point compared and the result is shown in fig3. Naïve Bayes has highest AUC of 0.71 compared with other classifiers.

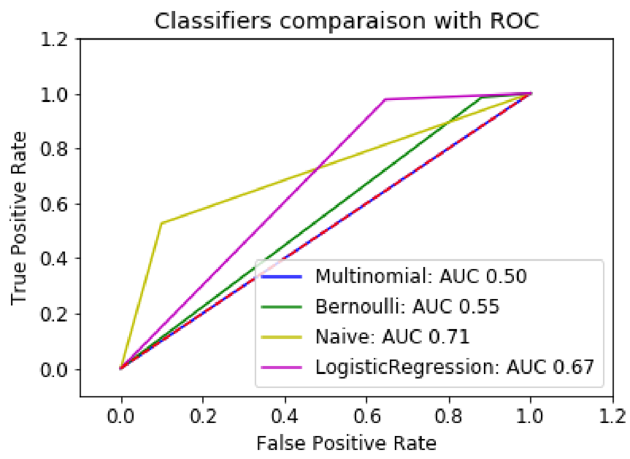

Figure 3. Classifiers comparison

Logistic Regression is having high precision of 0.92 , recall value is 0.93 and F1 Score value of 0.93 compared with other classifiers in Table 6. The reviews with empty ratings(NaN) are grouped and sentiment is predicted based on the review text. The review rating, reviews text, title, username, summary clean, words and classifiers result is in Table7. The review ratings are not entered by the user so the value is displayed as NaN.

Table 6. Precision and Recall of classifiers

classifier		precision	recall	f1-score	support
Naive:	positive	0.13	0.90	0.22	494
	negative	0.99	0.53	0.69	6438
	avg / total	**0.92**	**0.55**	**0.65**	**6932**
Multinomial:	positive	0.00	0.00	0.00	494
	negative	0.93	1.00	0.96	6438
	avg / total	**0.86**	**0.93**	**0.89**	**6932**
Bernoulli:	positive	0.38	0.12	0.18	494
	negative	0.94	0.99	0.96	6438
	avg / total	**0.90**	**0.92**	**0.90**	**6932**
	negative	0.95	0.98	0.97	6438
	avg / total	**0.92**	**0.93**	**0.93**	**6932**

The review text is taken for consideration and cleaned by CountVectorizer and TDIDF. The result of the cleaned text is shown in the below table under the column Summary_Clean.The features selected for sentiments is displayed as words. The classifiers results are 'pos', for positive and 'neg' for negative sentiments.*" The Kindle is my first e-ink reader. I own an iPad, an iPhone, and have owned a Windows-based phone in the past that I used as an ereader.My overall impression of the device is good.The good:I'd honestly rather read linear (read from page one to the end, one page at a time) fiction from it than a book, because I can't always get comfortable with a book. Hardcovers are sometimes a bit heavy, and paperbacks don't always lie open easily. The Kindle is incredibly light and thin. I can hold it in one hand easily. The page turn buttons are conveniently located. Page-turns aren't instant, but they're probably quicker than turning a physical page in a printed book (there are just a lot more page-turns unless you choose a small font). The contrast is better than other ereaders I've seen. There is zero eye strain in good light. My eyesight isn't* *the greatest and I like being able to increase the font size and read without glasses. I love being able to browse the Kindle store and read samples before deciding to purchase. The experimental browser is surprisingly usable, but isn't great. It is useful for browsing wikipedia and blogs. The biggest drawback to the browser is the awkward pointer navigation, using the 5-way pad. It syncs your furthest read page over the internet so you can pick up where you left off using your iPhone or iPad.The so-so:The kindle store could use more categories and sorting options. You can't sort by top rated, and there is no category for alternate histories, for example. Finding a very-specific type of fiction relies on keyword searches, which don't do a great job. The wifi sometimes doesn't connect before it times-out."* Is the review text in the first row of the below table. Even though the reviewer started with the positive note the opinion of the reviewer about the product is negative. The classifiers Naïve Bayes and Logistic regression is able to predict the sentiment correctly compared to the other classifiers.

Table 7. Prediction for reviews with empty ratings

	reviews.rating	reviews.text	reviews.title	reviews. username	Summary_ Clean	words	Naive	multi	Bill	log
0	NaN	The Kindle is my first e-ink reader. I own an ...	Worth the money. Not perfect, but very very go...	Jeffrey Stanley	the kindle is my first e ink reader i own an i...	[the, kindle, is, my, first, e, ink, reader, i...	neg	neg	neg	neg
1	NaN	I'm a first-time Kindle owner, so I have nothi...	I Wanted a Dedicated E-Reader, and That's What...	Matthew Coenen	i m a first time kindle owner so i have nothin...	[i, m, a, first, time, kindle, owner, so, i, h...	neg	neg	neg	neg
2	NaN	UPDATE NOVEMBER 2011:My review is now over a y...	Kindle vs. Nook (updated)	Ron Cronovich	update november my review is now over a year o...	[update, november, my, review, is, now, over, ...	neg	neg	neg	neg
3	NaN	I'm a first-time Kindle owner, so I have nothi...	I Wanted a Dedicated E-Reader, and That's What...	Matthew Coenen	i m a first time kindle owner so i have nothin...	[i, m, a, first, time, kindle, owner, so, i, h...	neg	neg	neg	neg
4	NaN	I woke up to a nice surprise this morning: a n...	Not the perfect do-it-all device, but very clo...	C. Tipton	i woke up to a nice surprise this morning a ne...	[i, woke, up, to, a, nice, surprise, this, mor...	neg	neg	neg	neg
5	NaN	The Kindle is my first e-ink reader. I own an ...	Worth the money. Not perfect, but very very go...	Jeffrey Stanley	the kindle is my first e ink reader i own an i...	[the, kindle, is, my, first, e, ink, reader, i...	neg	neg	neg	neg
6	NaN	UPDATE NOVEMBER 2011:br /br /My review is now ...	Kindle vs. Nook (updated)	Ron Cronovich	update november br br my review is now over a ...	[update, november, br, br, my, review, is, now...	neg	neg	neg	neg
7	NaN	I woke up to a nice surprise this morning: a n...	Not the perfect do-it-all device, but very clo...	C. Tipton	i woke up to a nice surprise this morning a ne...	[i, woke, up, to, a, nice, surprise, this, mor...	neg	neg	neg	neg
8	NaN	I use to hate to read but now that I have my K...	Great	D. Tatro	i use to hate to read but now that i have my k...	[i, use, to, hate, to, read, but, now, that, i...	neg	pos	pos	neg
9	NaN	All of them quit working. There's absolutely n...	I've had 3!	M. Lansford Kindle fave	all of them quit working there s absolutely no...	[all, of, them, quit, working, there, s, absol...	neg	pos	pos	neg

The model is tested with handwritten samples of text. The statement no.1, *"The product was fantastic"* predicted as positive and the probability of positive is 0.999 and negative is 0.000 which means 99% positive and 0% negative. *The statement* no.2*,"ohh gosh what is this useless!!"* is predicted as negative *and statement* no.3*, "product is not good"* also predicted as negative in Table 8.The results also verified manually and found correct. Our system found suitable for the open sentiments and results are predicted automatically.

Table 8. Sample text and predicted Result

Statement	Result
1	Sample estimated as POS: negative prob 0.000010, positive prob 0.999990
2	Sample estimated as NEG: negative prob 0.999993, positive prob 0.000007
3	Sample estimated as NEG: negative prob 0.996434, positive prob 0.003566

The sentiments extracted as features are displayed in word cloud using Python. The size of the word displayed in the Wordcloud is related to the count. Fig 4. Shows all the sentiment words extracted as features for opinion.

Figure 4. Word cloud of all sentiment words

The words used for predicting positive sentiments are shown in fig 5. and negative sentiment words in fig 6.

Figure 5. Word cloud of positive words

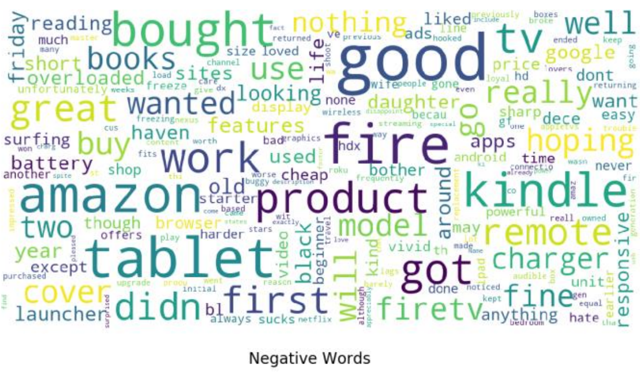

Figure 6. Word cloud of all negative words

CONCLUSION

Sentiment mining plays a very important role in business to understand the opinion of customers to improve the products Customer also depends on the opinion of others who has bought the products already. Reviews or feedback becomes the deciding factor for buy or sell a product. A rating of the products gives a speedy clarification to pact with the product. Most of reviews have lengthy comments without ratings. It is a time consuming factor to read all the reviews and come to a conclusion. Our model predicts the opinion from the reviews without ratings. It is observed that the logistic regression and Naïve Bayes predictions of opinions are much similar than the multinomial and Bernouli classifiers. In few cases logistic regression performance is better than the Naïve Bayes. In the future work we focus on the feature selection techniques to make the predictions of opinions natural.

REFERENCES

[1] Xing Fan, Justin Zhan, Sentiment analysis using product review data, Journal of Big Data 2015 2:5.

[2] www.amazon.com

[3] Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande, "Amazon Review Classification and Sentiment Analysis ",IJCSIT, Vol. 6 (6) , 2015, 5107-5110.

[4] Ahmad Kamal , "Subjectivity Classification using Machine Learning Techniques for Mining Feature - Opinion Pairs from Web Opinion Sources", International Journal of Computer Science Issues (IJCSI), Volume 10 Issue 5, 2013, pp 191-200.

[5] Bhumika Gupta , Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani , "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python", International Journal of Computer Applications (0975 – 8887) Volume 165 – No.9, May 2017.

[6] Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04.. Association for Computational Linguistics, Stroudsburg, PA, USA

[7] Rish, I, "An Empirical Study of the Naive Bayes Classifier", IJCAI 2001 Workshop on Empirical Methods in AI (2001).

[8] Evandro Costa, Rafael Ferreira, Paulo Brito, Ig Ibert Bittencourt, Olavo Holanda, Aydano Machado, Tarsis Marinho "A framework for building web mining applications in the world of blogs: A case study in product sentiment analysis", Expert Systems with Applications (39),pp.4813-4834,2012.