JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

Performance Evaluation of Machine Learning Models for Prediction of Sentiments in Movie Reviews

Dr. P. Sengottuvelan
Research Supervisor
Associate Professor in Computer Science
P.G Extension Centre Periyar University
Dharmapuri. Tamil Nadu, INDIA
sengottuvelan@gmail.com

I Anette Regina
Research Scholar
Associate Professor in Computer Science
Periyar University
Salem, Tamil Nadu, INDIA
anette1967cs@gmail.com

Abstract

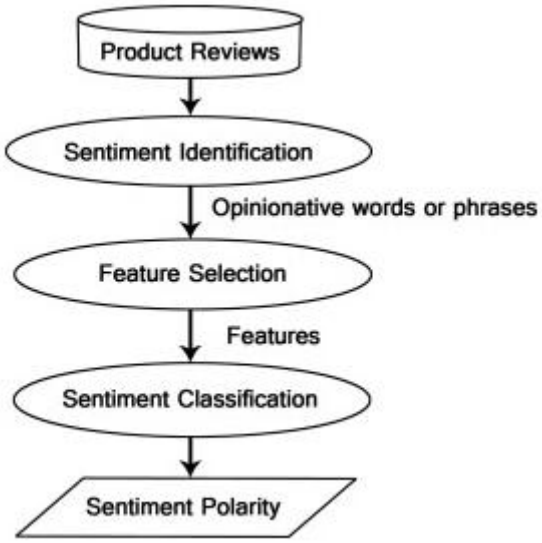
Sentiment Analysis is the most prominent branch of natural language processing. It deals with the text classification in order to determine the intention of the author of the text. The intention can be of admiration (positive) or criticism (Negative) type. This paper presents a comparison of results obtained by applying Naive Bayes (NB), Maximum Entropy, Random Forest, XGBoost, Logistic Regression and Support Vector Machine (SVM) classification algorithm. These algorithms are used to classify a sentimental review having either a positive review or negative review. The dataset considered for training and testing of model in this work is labelled based on polarity movie dataset and a comparison with results available in existing literature has been made for critical examination.

Keywords: *Sentiment Analysis, Naive Bayes (NB), Support Vector Machine (SVM), Classification, Polarity Movie Dataset*

Introduction

The domain Sentiment Analysis (SA) and Opinion mining (OM) are directly related to identifying the opinion of people, their attitude towards a particular scene or scenario and their emotions towards the same. The scene or scenario can be related to any event or topic. The people can give their reviews on any event at any time. Though some researchers say that both SA and OM are almost same, there are slight differences. Opinion Mining is the technique using which the opinion of people are extracted and analysed for a particular event where as Sentiment Analysis is the one which identifies the emotion or sentiment that is expressed by the text and then if required analyses the event. Hence by Sentiment Analysis one can identify the sentiment on a text or review and classify the polarity saying as either positive or negative. The process is as shown in the figure below.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020



The process of analysing the sentiments can be considered as a classification process. The classification can be done in three levels namely document level, sentence level and aspect level. Document Level is the very basic type of classification which considers the entire document for identifying whether it is a positive sentiment or negative sentiment document. Sentence level is the next level, where every sentence in the document is considered as the basic unit for classifying the polarity as either positive or negative. The major drawback of the first two levels are that, they are considered only in case of subjective scenarios. But these kinds of classification would not give a perfect classification of positive and negative sentiments. To increase the accuracy of the classification, we need to go to the Aspect Level which is the third level. Here every entity in the sentence is considered. For example, if the sentence is,

"The quality of the speaker in the Television is not good but the picture quality is best."

If the review is as specified above, then document level and sentence level SA would not give you a best classification. In such cases, Aspect level is used.

The research on sentiment analysis still has open issues which are not dealt with. The major issue with SA is the data set. The data sets that are currently available are mostly related to product reviews. These kinds of reviews are useful for business persons based on which they can take important decisions for further improvement of their product. SA can also be applied on other domains like stock, news, politics, etc. The social media plays a major role for these kinds of reviews. There are variety of applications and advancements in the domain

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

Sentiment Analysis. This paper focuses on discussing the various machine learning based methodologies used for classifying the sentiment of the movie reviews.

Related Work

An author Pang et.al. has given an approach for classifying the sentiments based on aspect level and categorized them into positive and negative polarity using machine learning algorithms namely Naive Bayes, Support Vector Machine and Maximum Entropy. They applied all these techniques by n-gram method. Another researcher tried to classify the sentiments using unsupervised learning algorithm. He tried to classify them by identifying the adjectives and adverbs in the sentence. An author used the scoring methods to identify whether the reviews that are structured are positive reviews or negative reviews. Pang and Lee tried to extract the reviews and labelled them as either subjective or objective sentence. Then they used the minimum cut formulation methodology for exploring the methods. Their technique prevented the features which were misleading the classification. Whitelaw et.al has proposed a sentiment classification technique based on the analysis and extraction of appraisals. Another researcher had proposed a semi-supervised method using the sampling methodology. They have tried to overcome the issue of imbalance. A variance mean based methodology is proposed by Wang and Wang which is a feature based filtering method. This method minimizes the number of features required for text based classification. The method tried to extract only the best features and classify accordingly. Due to this the time required for computation also was reduced.

Methodology

In general, two kinds of approaches are used for analysing the sentiments or classifying the sentiments based on polarity namely binary sentiment classification and multi-class sentiment classification. IN case of binary, every sentence or document or aspect is classified as either positive or negative. But in case of multi-class, there can be more than two that is, strongly positive, positive, strongly negative, negative and neutral. Most of the researchers concentrate on binary classification. In this research, the focus is towards binary classification based on machine learning approaches. The use case considered for binary classification is the reviews given by people on a movie or in simple terms we can say as movie reviews. The algorithms or methodologies used are discussed below.

Naïve Baye's Classifier (NB)

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

This is a probabilistic based classifier that uses the mathematically proven Baye's Theorem. The major assumption made is that all features are independent. The major advantage of the algorithm is that the data required for preparing the model is very less. The methodology builds a feature vector using only the variance of the features. The conditional probability for each review "R" can be calculated using the equation given below.

$$P(c \mid R) = \frac{P(R \mid c) * P(c)}{P(R)}$$

Support Vector Machine (SVM)

This is a methodology which could be used as a binary classifier and is non-probabilistic in nature. We have modelled the classifier so that each review is considered as a vector and is placed as a data point on the space. The most important point in this classifier is to identify the hyperplane which can help us to classify the reviews. The hyperplane will separate all the data so that there is no error and the closest points are moved apart from the plane. The hyperplane can be identified for a document "d" using the below equation.

$$\vec{w} = \sum_j \alpha_j c_j \vec{d}_j, \alpha_j \geq 0$$

The performance is analysed using the confusion matrix which is generated as follows. The matrix represents the relation between the rightly predicted sentiments and the wrongly predicted ones. Four values are tabulated in the confusion matrix namely True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). Using these values different measures could be analysed namely Precision, Recall, F-measure and accuracy as follows.

Precision: It gives the exactness of the classifier. It is the ratio of number of correctly predicted positive reviews to the total number of reviews predicted as positive.

$$precision = \frac{TP}{TP+FP}$$

Recall: It measures the completeness of the classifier. It is the ratio of number of correctly predicted positive reviews to the actual number of positive reviews present in the corpus.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

$$Recall = \frac{TP}{TP+FN}$$

F-measure: It is the harmonic mean of precision and recall. F-measure can have best value as 1 and worst value as 0. The formula for calculating F-measure is presented as:

$$F - Measure = \frac{2*Precision*Recall}{Precision+Recall}$$

Accuracy: It is one of the most common performance evaluation parameter and it is calculated as the ratio of number of correctly predicted reviews to the number of total number of reviews present in the corpus. The formula for calculating accuracy is given as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Results and Discussion

Data Description

There are various publicly available websites like Rotten Tomatoes, IMDB and Flixter in which the people can share their opinion about a movie based on their thoughts. We have used the "Large Movie Review Dataset" which is usually referred as IMDB Dataset. For the purpose of analysing the sentiment and classifying the opinion as either positive or negative, we use 50,000 records from the entire dataset. Out of these records, we use 25,000 records for training and 25,000 records for testing the various machine learning algorithms. The figure 1 below represents the review of the length of words in the data set.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

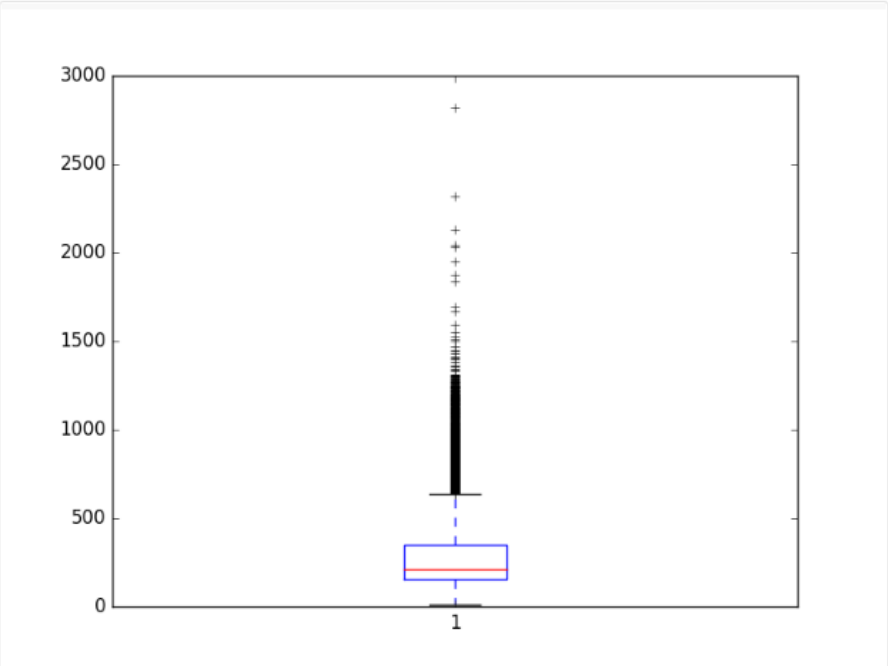


Figure 1: Length of Words in the Data set

Data Pre-processing

The first phase of the sentiment analysis process is to prepare the data for evaluation. The dataset contains only raw text with label saying either positive or negative. The steps involved for preparing the data are as follows.

1. Tokenization and Segmentation
2. Noise Removal
3. Lemmatization and Normalization
4. Vectorising

Tokenization and Segmentation

We use the library NLTK from python for performing this step. This phase subdivides the single sentence into few sub-strings called "Tokens". For example, consider the review as given below

"One would expect the movie to be commercial"

The output of the phase would be

['One', 'would', 'expect', 'the', 'movie', 'to', 'be', 'commercial']

Noise Removal

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

There is no specific package in python for perfectly removing the stop words. We designed a file which has few set of stop words and whichever matches the token, we tried to remove considering them as noise. Few examples of words that are removed are as shown in the table 1.

Table 1: Noise in the data

Emotions	:), : (
Hyperlinks	https:// , @
Punctuations	!, :, ;, …
Numeric	0-9
Articles and prepositions	a, an, the, before, and, though

Stemming and Normalization

This phase identifies the similar words in the data or the input test. Few examples of similar words are "movies","movie", "film". Hence stemming is the process of identifying the similar meaning words and group them. A very widely used algorithm for performing this stemming process is "Skip-Gram with Negative Sampling"

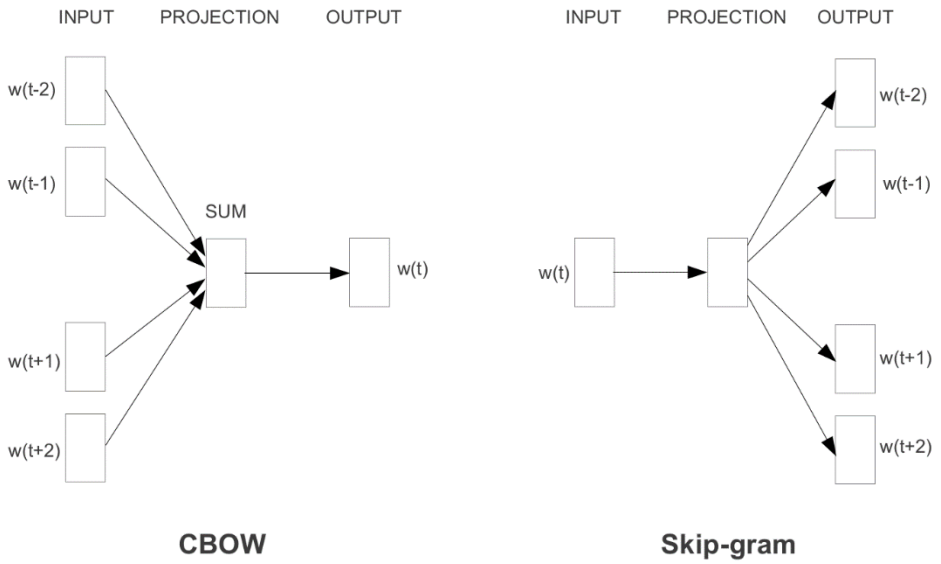


Figure 2. Word to Vector Conversion

Vectorization

There are three kinds of feature selection mechanisms namely Bag of words, TF-IDF and Word Embedding. Bag of Words is a very straight forward method in which a file consists of all possible keywords which would be considered as positive and then can be matched. TF-IDF is the abbreviation of Term Frequency-Inverse Document Frequency. This method

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

calculates the frequency of occurrence of the words in the set. Frequently occurring words are given lesser weights and rare words will be given higher weights. Word embedding is a very popular method which converts the words into a numeric value called as vector using English dictionary. The final output of this stage is the feature vector.

Machine Learning Models for analysing the sentiment

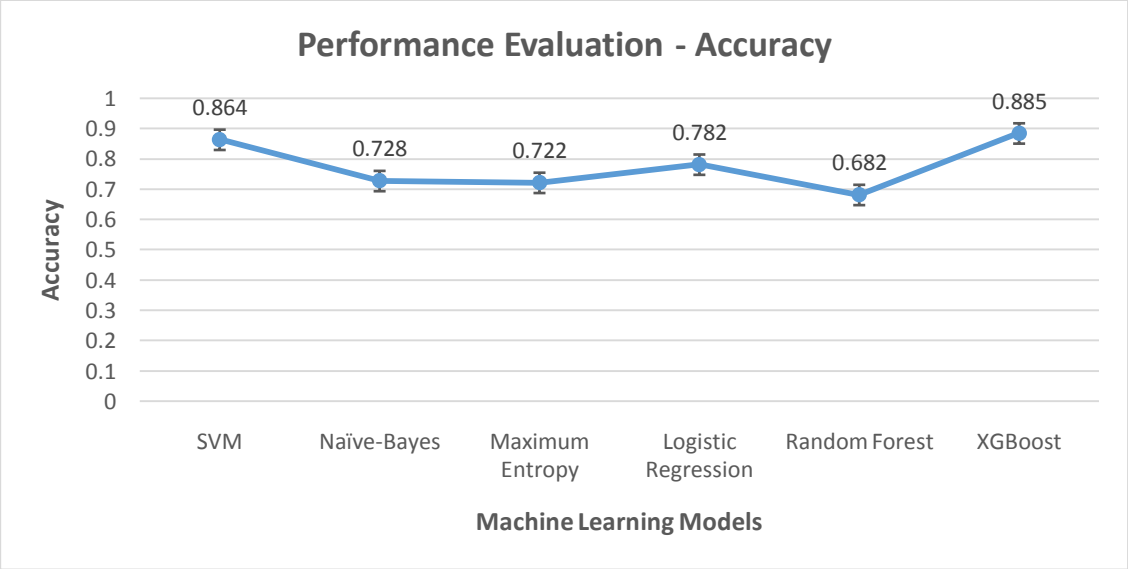
The polarity of the review is classified into either "positive" or "negative" using widely used Machine Learning models namely Support Vector Machine(SVM), Naïve-Bayes Classifier, Maximum Entropy, Logistic Regression, Random Forest and XGBoost.The performance of the models are evaluated using the metrics F-measure, Precision, Recall and Accuracy. The obtained metrics for each Machine Learning modes are tabulated in the table 2 below.

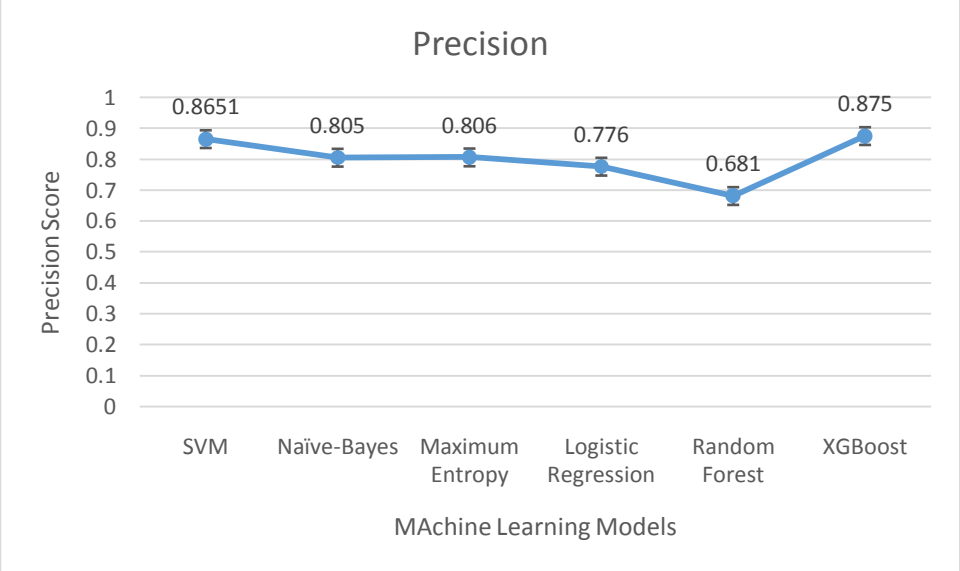
Table 2: Performance analysis of Machine Learning Models

ML Models	Accuracy	Precision	Recall	F-measure
SVM	0.864	0.8651	0.864	0.8638
Naïve-Bayes	0.728	0.805	0.728	0.7095
Maximum Entropy	0.722	0.806	0.722	0.7015
Logistic Regression	0.782	0.776	0.771	0.7595
Random Forest	0.682	0.681	0.672	0.6819
XGBoost	0.885	0.875	0.872	0.8713

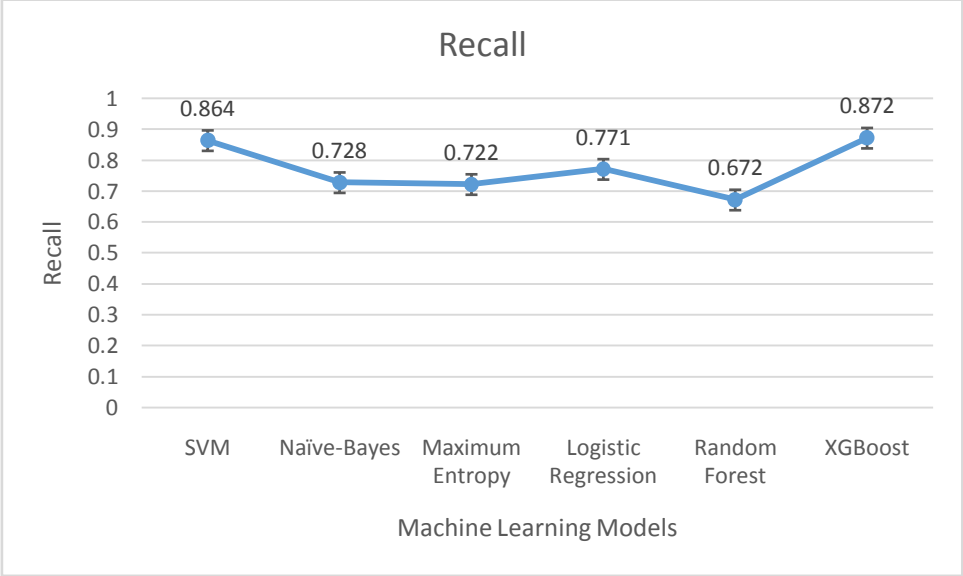
The performance metrics are plotted comparing all the machine learning models in the figure 1 to 4.

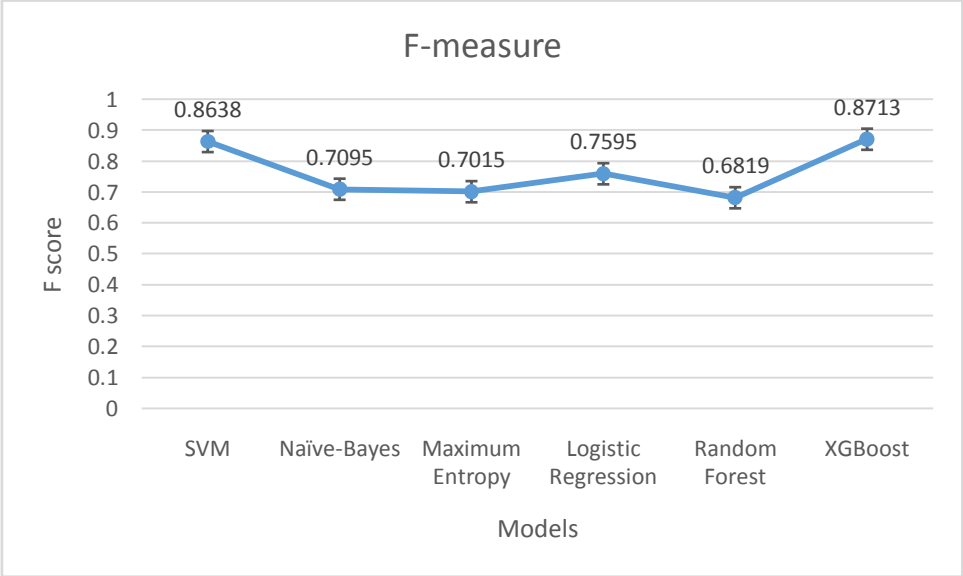
ISSN- 2394-5125 VOL 7, ISSUE 08, 2020





ISSN- 2394-5125 VOL 7, ISSUE 08, 2020





From the results it is evident that the XGBoost Model outperforms all the other available models. The results shown are for binary classification. The same process can be performed for multi-class classification as well.

Conclusion

In this study, an attempt has been made to classify sentiment analysis for movie reviews using machine learning techniques. Different algorithms namely Naive Bayes (NB), Logistic Regression, Maximum Entropy, Random Forest, XGBoost, and Support Vector Machine (SVM) are implemented. These algorithms have also been implemented earlier by different researchers and results of all versions of

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

implementation have been compared. It is observed that XGBoost classifier outperforms every other classifier in predicting the sentiment of a review.

References

[1]TsytsarauMikalai, Palpanas Themis. Survey on mining subjective data on the web. Data MinKnowl Discov2012;24:478–514.

[2] Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarityin phrase-level sentiment analysis. In: Proceedings of HLT/EMNLP; 2005.

[3]Liu B. Sentiment analysis and opinion mining. Synth LectHuman Lang Technol 2012.

[4]Yu Liang-Chih, Wu Jheng-Long, Chang Pei-Chann, Chu Hsu-an-Shou. Using a contextual entropy model to expand emotionwords and their intensity for the sentiment classification of stockmarket news. Knowl-Based Syst 2013;41:89–97.

[5] Michael Hagenau, Michael Liebmann, Dirk Neumann. Auto-mated news reading: stock price prediction based on financialnews using context-capturing features. DecisSuppSyst; 2013.

[6]Tao Xu, Peng Qinke, Cheng Yinzhao. Identifying the semanticorientation of terms using S-HAL for sentiment analysis. Knowl-Based Syst 2012;35:279–89.

[7]Maks Isa, VossenPiek. A lexicon model for deep sentimentanalysis and opinion mining applications. Decis Support Syst2012;53:680–8.

[8]Pang B, Lee L. Opinion mining and sentiment analysis. FoundTrends Inform Retriev 2008;2:1–135.

[9]Cambria E, Schuller B, Xia Y, Havasi C. New avenues inopinion mining and sentiment analysis. IEEE Intel Syst2013;28:15–21.

[10]Feldman R. Techniques and applications for sentiment analysis.Commun ACM 2013;56:82–9.

[11]Montoyo Andre´s , Martı´nez-Barco Patricio, Balahur Alexandra.Subjectivity and sentiment analysis: an overview of the currentstate of the area and envisaged developments. Decis Support Syst2012;53:675–9.

[12]QiuGuang, He Xiaofei, Zhang Feng, Shi Yuan, Bu Jiajun, ChenChun. DASA: dissatisfaction-oriented advertising based onsentiment analysis. Expert SystAppl 2010;37:6182–91.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

[13]Lu Cheng-Yu, Lin Shian-Hua, Liu Jen-Chang, Cruz-LaraSamuel, Hong Jen-Shin. Automatic event-level textual emotionsensing using mutual action histogram between entities. ExpertSystAppl 2010;37:1643–53.

[14] Neviarouskaya Alena, Prendinger Helmut, Ishizuka Mitsuru.Recognition of Affect, Judgment, and Appreciation in Text. In:Proceedings of the 23rd international conference on computational linguistics (Coling 2010), Beijing; 2010. p. 806–14.

[15]Bai X. Predicting consumer sentiments from online text. DecisSupport Syst 2011;50:732–42.

[16]Zhao Yan-Yan, Qin Bing, Liu Ting. Integrating intra- and inter-document evidences for improving sentence sentiment classification. ActaAutomaticaSinica 2010;36(October'10).

[17]Yi Hu, Li Wenjie. Document sentiment classification by exploring description model of topical terms. Comput Speech Lang2011;25:386–403.

[18]Cao Qing, DuanWenjing, GanQiwei. Exploring determinants ofvoting for the ''helpfulness'' of online user reviews: a text miningapproach. Decis Support Syst 2011;50:511–21.

[19]He Yulan, Zhou Deyu. Self-training from labeled features forsentiment analysis. Inf Process Manage 2011;47:606–16.

[20]Tan Songbo, Wu Qiong. A random walk algorithm forautomatic construction of domain-oriented sentiment lexicon.Expert SystAppl 2011:12094–100.

[21]Tan Songbo, Wang Yuefen. Weighted SCL model for adaptationof sentiment classification. Expert SystAppl 2011;38:10524–31.

[22]Qiong Wu, Tan Songbo. A two-stage framework for cross-domain sentiment classification. Expert Syst Appl2011;38:14269–75.

[23] Jiao Jian, Zhou Yanquan. Sentiment Polarity Analysis basedmulti-dictionary. In: Presented at the 2011 International Conference on Physics Science and Technology (ICPST'11); 2011.

[24] LambovDinko, PaisSebastia˜o, Dias Ga˜el . Merged agreementalgorithms for domain independent sentiment analysis. In:Presented at the Pacific Association for, Computational Linguistics (PACLING'11); 2011.

[25]Xu Kaiquan, Liao Stephen Shaoyi, Li Jiexun, Song Yuxia.Mining comparative opinions from customer reviews for competitive intelligence. Decis Support Syst 2011;50:743–54.

[26]Chin Chen Chien, Tseng You-De. Quality evaluation of productreviews using an information quality framework. Decis SupportSyst 2011;50:755–68.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

[27]Fan Teng-Kai, Chang Chia-Hui. Blogger-centric contextualadvertising. Expert SystAppl 2011;38:1777–88.

[28] Zhou L, Li B, Gao W, Wei Z, Wong K. Unsupervised discoveryof discourse relations for eliminating intra-sentence polarityambiguities. In: Presented at the 2001 conference on EmpiricalMethods in Natural Language Processing (EMNLP'11); 2011.

[29] Heerschop B, Goossen F, Hogenboom A, Frasincar F, KaymakU, de Jong F. Polarity Analysis of Texts using DiscourseStructure. In: Presented at the 20th ACM Conference onInformation and Knowledge Management (CIKM'11); 2011.

[30] Zirn C, Niepert M, Stuckenschmidt H, Strube M. Fine-grainedsentiment analysis with structural features. In: Presented at the5th International Joint Conference on Natural Language Processing (IJCNLP'11); 2011.

[31]Hu Nan, Bose Indranil, KohNoi Sian, Liu Ling. ''Manipulationof online reviews: an analysis of ratings, readability, andsentiments''. Decis Support Syst 2012;52:674–84.

[32]Gupta Sunil Kumar, PhungDinh, Adams Brett, VenkateshSvetha. Regularized nonnegative shared subspace learning. DataMinKnowlDiscov 2012;26:57–97.

[33]Duric Adnan, Song Fei. Feature selection for sentiment analysisbased on content and syntax models. Decis Support Syst2012;53:704–11