Web Application For Sentimental Analysis Using Machine Learning

Prof. Ms. Pranita P. Deshmukh[1], Prof. Ms. Rutuja A. Gulhane[2]

Prof. Ms. Rupali Meshram[3] , Prof. Ms. Rani Lande[4]

Assistant Professor, Dept. Of Computer Sci. & Engg., PRMIT&R Badnera, Amravati, India [1]

Assistant Professor, Dept. Of Computer Sci. & Engg., PRMIT&R Badnera, Amravati, India [2]

Assistant Professor, Dept. Of Computer Sci. & Engg., PRMIT&R Badnera, Amravati, India [3]

Assistant Professor, Dept. Of Computer Sci. & Engg., PRMCEAM Badnera, Amravati, India [4]

***ABSTRACT*: *Sentiment Analysis is the 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 (+ve) or criticism (-ve) type. This paper presents a comparison of results obtained by applying Naive Bayes (NB) and Support Vector Machine (SVM) classification algorithm. These algorithms are used to classify a sentimental review having either a positive review or negative review. Sentiment analysis is now in focus of companies to extract information from customer reviews. Usually the analysis classification is positive, negative and neutral. In this research we focus on the reviews for electronic products. The demographic and technical expertise level of consumers and reviewers are diverse hence there is difference in the way they review a product. Some review contains technical solutions, improvised ways of tackling problems, frustrations, joy etc. This difference in review calls for a wider classification scheme to contain these differences. We thereby introduced a five classification scheme namely positive, negative, advice, no sentiment and neutral at the sentence. We crawled data from amazon.com and used open source natural language processing tools to get the sentiment out of the review*.**

1. INTRODUCTION

The ever increasing use the Internet for e-commerce has produced vast large consumer generated contents such as reviews. It is becoming imperative for organizations to collect these data and useful insight can be mined from the reviews. As prices of storage space is decreasing it serves as the right step to collect review from web for analysis to know what users are saying about a particular product. Organization can use these user feedbacks to make their upcoming products better or improve their customer service.

Sentiment mainly refers to feelings, emotions, opinion or attitude1. With the rapid increase of world wide web, people often express their sentiments over internet through social media, blogs, rating and reviews. Due to this increase in the textual data, there is a need to analyze the concept of expressing sentiments and calculate the insights for exploring business. Business owners and advertising companies often employ sentiment analysis to discover new business strategies and advertising campaign.

Machine leaning algorithms are very often helpful to classify and predict whether a document represents positive or negative sentiment. Machine learning is categorized in two types known as supervised and unsupervised machine learning algorithms. Supervised algorithm uses a labeled dataset where each document of training set is labelled with appropriate sentiment. Whereas, unsupervised learning include unlabeled dataset where text is not labeled with appropriate sentiment. This study mainly concerns with supervised learning techniques on a labeled dataset.

Rating of user depends on the overall performance and experience. Some reviews are very short and some are very verbose covering hundred plus lines. The more detailed review shows the technical knowledge of the reviewer. As there is not a strict language system in reviewing, a lot of users use slangs and texting words and emoticon.

Due to the popularity of reviews there has been a lot of Sentiment Analysis and text mining applied to reviews such as movies. These techniques usually places into three categories namely positive, negative and neutral. The positive is favourable experience, negative for bad experience and neutral is no sentiment. However analyzing reviews on the sentence level, the sentences can be placed in more than three categories. Users with more technical knowledge gives more information such as how to circumvent problems face, recommendation to improve a product. Some sentences just give background information of where the product were bought.

In this paper we will focus on consumables. We will apply a three classification scheme namely recommendation or advice, negative, positive, neutral to these sentences. Our primary data is from Television reviews of LG and Samsung.

2. SENTIMENT ANALYSIS

Sentiment analysis is computational study of emotions, opinions and mainly the sentiment expressed in the text by user. Sentiment analysis is a challenging task due to many challenges
which are associated while processing natural language. Any sentiment analysis system needs
first to extract feature i.e. sentimental words or phrases from the given text and then using
suitable text classifier overall sentiment associated with the text is extracted.

2.1 TYPES OF SENTIMENTAL ANALYSIS

Sentiment analysis is usually implemented on three levels namely sentence level, document level and aspect level.

Document Level sentiment classification aims at classifying the entire document or topic as positive or negative.

Sentence level sentiment classification considers the polarity of individual sentence of a document .

Aspect level sentiment classification first identifies the different aspects of a corpus and then for each document, the polarities calculated with respect to obtained aspects.

2.2. Challenges for sentiment analysis

1. Contextual Information

Identifying the context of the text becomes an important challenge to address in SA.

Behaviour/use of the word changes in a great aspect based on the context.

Ex-1 The journey was long.

Ex-2 Seminar was long.

Ex-3 Battery life of Nexus 5 is long.In all the above 3 examples, meaning of long is same it indicates the duration or passage of time.In Ex-1 and Ex-2 “long” indicates bored hence a Negative expression whereas in Ex 3 “long”indicates efficiency hence a Positive expression. In Ex 3 “long” indicates efficiency hence a Positive expression.

2. Sarcasm Detection

Sarcasm involves statement and a remark which is usually indirect taunt towards any object or an appraisal in a negative way. Detecting sarcasm is a tough task for humans and equally harder for machine. Some examples of sarcasm: Ex- Amazing presentation by Mr. X, I won't ever attend such presentation again.

3. Word Sense Disambiguation

Word sense disambiguation (WSD) is the problem of determining in which sense a word having a number of distinct senses is used in a given sentence. The same word can have multiple meanings, and based on the sense of its usage the polarity of the word also changes. For example, the word "cold" has several senses and may refer to a disease, a temperature sensation,or an environmental condition. The specific sense intended is determined by the textual context in which an instance of the ambiguous word appears. In "I am taking aspirin for my cold" the disease sense is intended, in "Let's go inside, I'm cold" the temperature sensation sense is meant,while "It's cold today, only 2 degrees", implies the environmental condition sense.

4. Word Order:

Word order plays a vital role in deciding the polarity of a text, in the text same set of words with slight variations and changes in the word order affect the polarity aspect. For example "X is efficient than Y" conveys the exact opposite sentiment from "Y is efficient than X".

5. Identify subjective portions of text:

The same word can be treated as subjective in one context and objective in some other. This makes it difficult to identify the subjective (sentiment-bearing) portions of text. Consider following examples:

Ex-1 the language of the author was very crude.

Ex-2 Crude oil is extracted from the sea beds.

6. Indirect negation of sentiment

Negation of sentiment is assigned to words like no, not, never, etc. But there are certain words tend to reverse the sentiment polarity implicitly. For example the sentence, "This movie avoids all predictable and boring drama found in most of the bollywood movies." The negative sentiment associated with words predictable and boring is reversed by associating word avoid with those words.

7. Entity Recognition

Same entity is not seen for all the texts in a document. When multiple entities are being mentioned about in a single document, the overall document polarity does not make much sense.We need to separate out the text about a particular entity and then analyze its sentiment. Ex- I hate Heat, but I like ICE.

2.3. Features for sentiment analysis

Converting a piece of text to a feature vector is the basic step in any data driven approach to SA.

Some commonly used features used in Sentiment Analysis are:

1.Term Presence vs. Term Frequency:

Term frequency is mostly used in different Information Retrieval (IR) and Text Classification

tasks. But Pang-Lee found that term presence is more important to Sentiment analysis than

term frequency as for term presence binary-valued feature vectors are used, and those vectors the entries merely indicate whether a term occurs indicated as value 1 or do not occur indicated by value 0. Polarity based classification is very useful for SA ,as overall sentiment may not usually be highlighted through repeated use of the same term an also occurrences of rare word can have more sentimental value compared to frequently repeated word.

2.Term Position:

More sentimental value is given to certain words based on their position in a sentence or a

document. Generally words appearing in the 1st few sentences and last few sentences in a text are given more weightage than those appearing elsewhere in the document.

3.Parts of speech (POS):

Part-of-speech (POS) information is commonly exploited in sentiment analysis which deals with finding adjectives, adverbs in the text as they are important indicators of sentiment in a given text. Amongst all parts of speech like nouns, verbs, adjectives, etc most important POS is considered as adjective to represent sentimental feature or simply subjective text. Adverbs when occurs along adjectives improves the probability of finding exact sentiment of a given text.

4.Topic-Oriented Features:

Interactions between topic and sentiment play an important role in extracting sentiment from the content. For example, in a hypothetical article on Rebook, the sentences “Reebok reports that profits rose” and “Target reports that profits rose” could indicate completely different types of news regarding the subject of the document. Topic information can also be incorporated into features set as they also help in specifying sentiment.

5.Opinion words and phrases:

There are words like; good or bad, like or hate, happy or sad which are mostly used to express opinions towards an object. Also sometimes there are phrases which expresses opinions without using an of the opinion words in the phrase. For example: It cost me my full month salary.

3. SYSTEM ARCHITECTURE

In this section, we describe system architecture and steps in the data life cycle for the sentimental analysis. Figure 1 shows data source (amazon.com), storage (MySQL), text analysis and mining, and visualization of the result on the Web.

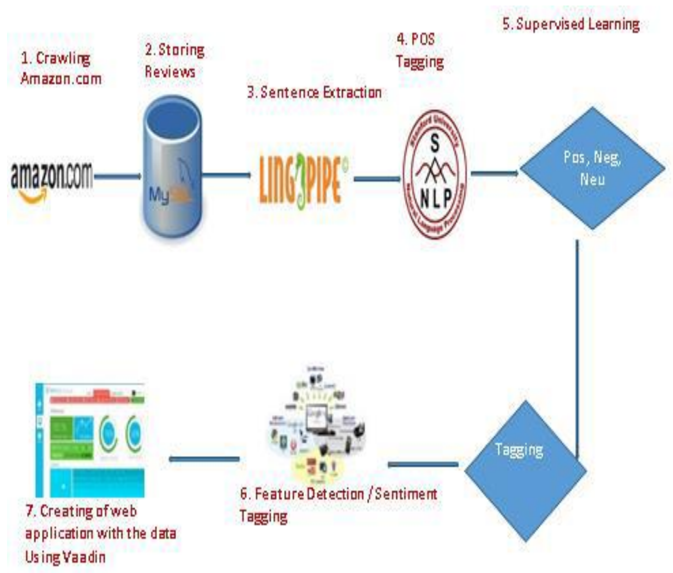


Figure 1. System Architecture for the Sentimental Analysis

In the next section, we formally propose the five classification scheme, feature detection, and geotagging the reviews.

3.1. Classification Scheme

Online reviews are usually classified by the reviewer so we will rather classify the sentence level. A 5 star review can also have some sentences which expresses a bad experience using a product. By reading through thousands of reviews some patterns of sentences emerged rather the three broad classification system. Consider the following sentences:

1. This TV is amazing and I'm glad I got it.
2. You will need a flat surface that is about 27 inches wide for this 32 inch TV.
3. The picture is excellent, however, the volume seems a bit muffled.
4. I was looking at a $600 Samsung with 120 Hz refresh rate, then I found this one.
5. The television cuts on, cuts off, randomly.
6. The machine will still rotate your clothes every so often to avoid wrinkles.

The above sentences show that reviews can be more than positive and negative and neutral. Hence we introduce a five class scheme namely positive, negative, advice, neutral and no sentiment. The first and sixth sentence will be placed in the positive category as they clearly state a positive feeling, the second would be place as an advice as it shows what surface the TV should be placed on. The third sentence has both positive and negative sentiment hence it would fall in the neutral category. The fourth would be placed in no sentiment category as there is no mood attached to it. Finally the fifth would be placed in the negative category as the reviewer wrote about a negative experience. Once we have our classification scheme set up we will proceed to the data collection.

(1) **Data Collection.**

We created a crawler to get the review from amazon.com using Jsoup API [19]. We collected data from the electronic section. We used the television, washing machines, refrigerators, notebooks and monitors products from Samsung and LG as our primary source of data. The information collected were the review text and the location of the reviewer. The crawled data was put in a MySQL database for later use.

(2) **Sentence Extraction**

As our focus was on the sentence level we had to split all the reviews into sentences before going to the next phase. We created a mentioned table for this.

(3) **Supervised Learning.**

Reviewers are not under any obligation to adhere to strict grammar rules hence the reviews sentences are unstructured. Many use emoticons, slangs or bad grammar. In one review the reviewer was so enthused about the product that she wrote "'Very goooood ~It made my husband happy.'" There were numerous spelling errors. This compelled us not to use the grammar context sentiment analysis and the SENTIWORDNET [11] approach. Also is very had to detect product reviews using grammar context as different products have different criteria to be described as good. For example a quiet washing machine or notebook is a positive sentiment whiles a quite TV or audio system is a negative sentiment. Additionally there are schemes to detected negative and positive but no scheme to detect recommendations are advices. These reasons compelled us to opt for a machine learning algorithm. We used LingPipe's Classifier API to train our classifier.

3.2. Feature Detection

There are many researches on extracting the topics of a review. In literature [13], Titov *et a l.*, formulated a statistical model which is able to discover corresponding topics in text and extract textual evidence from reviews. In literature [4], Zhai, Z., *et al.*, have developed a semi-supervised technique to identify clusters of features, *i.e.*, sets of synonyms that are likely to refer to the same product features. Their technique needs a fixed number of clusters, with a starting set of features. Our technique is a little similar of Zhai, Z., *et al*. In our approach we started with a list of product features. We used WordNet to get the Synset from this features to get more words. For each extracted sentence we converted it into lowercase before we used Stanford Part-Of-Speech (POS) tagger. Stanford POS tagger detects uppercased words as a noun, so in the case where a reviewer is shouting by typing all sentences in capital cases, the tagger might see all the words as nouns. The tagged nouns are looked up in the feature list created for validation. We also incorporated compound nouns detection, which is usually a feature in a feature. In case a compound noun is detected and a part can be found in the feature list it automatically detected as a feature.

3.3. Web Application

Having data in the database is not enough to get meaning from collection. To complete the process, visualization of the data is needed. There have been several attempts to do this. Liu *et al.*, [6] proposed a framework for analyzing and comparing consumer opinions of competing products where users can see the performance of competition products. Oelke *et al.*, [9] proposed a scalable alternative in order to aggregate large numbers of products and features, clustering similar users. Similarly, the system by Miao *et al.*, [8] visualizes the sentiment expressed in product reviews over time. Positive and negative opinions were aggregated over time and displayed with different charts. Wu *et al.*, [15] proposed the OpinionSeer, for reviews where uncertain sentiment through time is visually represented and aggregated. In this paper we present an application to visualize the aggregated data.

4. SENTIMENT CLASSIFICATION TECHNIQUE

Sentiment classification is a task under Sentiment Analysis (SA) that tags text as positive,negative or neutral automatically. Thus, a sentiment classifier tags the sentence 'the movie is entertaining and totally worth your money!' in a movie review as positive with respect to the movie. On the other hand, a sentence 'The movie is so boring that I was dozing away through the second half.' is labelled as negative. Finally, 'The movie is directed by Nolan' is labelled as neutral .There is main techniques for sentiment classification: machine learning based.

4.1.Machine Learning Approach

The machine learning method uses several learning algorithms to determine the sentiment by training on a known dataset. Training and a test set are the two document sets which are mostly needed for machine learning based techniques. Training set is used by classifier to understand the different characteristics associated with documents, and to check the overall performance of the classifier test set are used.

4.1.1.Supervised learning

The supervised learning methods depend on the existence of labelled training documents. Supervised learning process: two Steps; Learning (training): Learn a model using the training data testing: Test the model using unseen test data to assess the model accuracy. There are different types of supervised classifiers like: Rule-based Classifiers, Decision Tree Classifiers, Linear Classifiers and Probabilistic Classifiers.

4.1.2. Unsupervised learning

The unsupervised learning include unlabeled dataset where text is not labeled with appropriate sentiments2. This study mainly concerns with supervised learning techniques on a labeled dataset.

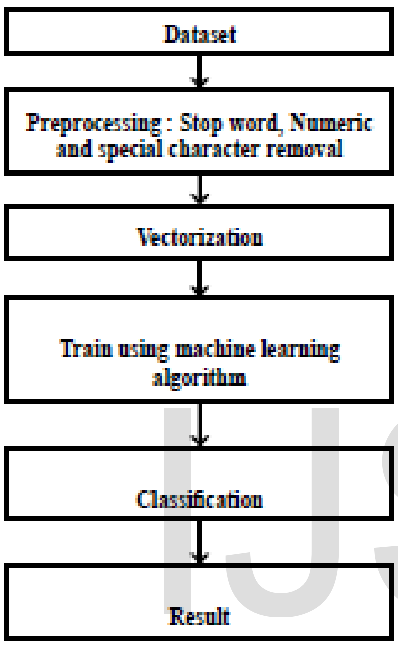
1. Naive Bayes Classifier (NB)

Naive Bayes classification model computes the posterior probability of a class is computed in Naive Bayes Classifier which is based on the way words are distributed in the particular document. The positions of the word in the document are not considered for classification in this model as it uses BOWs feature extraction technique. Bayes Theorem is used to predict the probability where given feature set belongs to a particular label of the content.

P(label_features)=(P(label)*P(features|label))/(P(features)) P (label) signifies the prior probability of a label. P (features | label) signifies the prior probability that a particular feature set is being classified as a label. P (features) specifies the prior probability

that a given feature set has occurred in the process. On basis of Naive assumption i.e. all features are independent; the equation can be rewritten as:
P(label_features)=(P(label)*P(f1_label)*…. *P(fn_label))/(P(features))

5. PROPOSED WORK



5.1. Steps Followed For Classification:

Step 1. The polarity movie review dataset is considered for analysis which consist of 1000 positive and 1000 negative labeled reviews. For each review a separate text file is maintained.

Step 2. The reviews contain a large amount of vague information which need to be eliminated. In preprocessing step, firstly, all the special characters used like (!@) and the unnecessary blank spaces are removed. It is observed that reviewers often repeat a particular character of a word to give more emphasis to an expression or to make the review trendy16. Words like *wooowwwwwww*, *oohhhhhh* falls in this category. The repetition of characters are also eliminated in this step.

Most of the words that do not contribute to any sentiment used in English language are termed as stopwords. So, second step in preprocessing involves the removal of all the stopwords of English language.

Step 3. After cleaning the dataset in step 3, features can be extracted from it. The features are tokenized word of a review. These words need to be converted to numerical vectors so that each review can be represented in the form of numerical data. The vectorization of features are done using the following two methods.

• **CountVectorizer**: It transforms the review to token count matrix. First, it tokenizes the review and according
to number of occurrence of each token, a sparse matrix is created.

– Calculation of CountVectorizerMatrix: suppose we have three different documents containing following
sentences.
”Movie is great”.
”Movie is Awful”.
”Movie is fine”.

Matrix generated of size 3*5 because we have 3 documents and 5 distinct features. The matrix will
look like given in table 2.

	Feature1	Feature 2	Feature 3	Feature 4	Feature 5
Sentence 1	1	1	1	0	0
Sentence 2	1	1	0	1	0
Sentence 3	1	1	0	0	1

Table 2: Matrix generated under CountVectorizer Scheme

Each 1 in a row corresponds to presence of a feature and 0 represents absence of a feature from particular document.

• **TF-IDF**: Its value represents the importance of a word to a document in a corpus. TF-IDF value is proportional to the frequency of a word in a document.

– Calculation of TF-IDF value : suppose a movie review contain 100 words wherein the word *Awesome* appears 5 times. The term frequency (i.e., TF) for *Awesome* then (5 / 100) = 0.05. Again, suppose there are 1 million reviews in the corpus and the word *Awesome* appears 1000 times in whole corpus.

Then, the inverse document frequency (i.e., IDF) is calculated as log(1,000,000/1,000) = 3. Thus, the TF-IDF value is calculated as: 0.05 * 3 = 0.15.

Step 4. The numeric vectors can be given as input to the classification algorithm.Algorithm used is as follows:

- Naive Bayes (NB) algorithm:

Initially, the dataset was not divided between testing and training subsets. So, k-fold cross validation technique is used, the number of folds used are 10.

Step 5. After training of model, confusion matrix is generated which shows the number of positive and negative reviews that are correctly predicted and number of positive and negative reviews that are wrongly predicted. For each fold, prediction accuracy is calculated based on this confusion matrix and final accuracy is given by taking the mean of all the individual accuracies of 10 folds. However, individual accuracy of a particular fold can be much higher than the mean of all accuracies.

Step 6. For each model, values of precision, recall and F-measure as performance evaluation parameters are found out.

The confusion matrix and a table containing performance evaluation parameter is generated. Finally, these result are obtained with proposed approach.

5.2. Implementation

The implementation of above mentioned algorithms are carried out on Polarity movie review dataset. K-fold cross validation algorithm is implemented where single fold is considered for testing and remaining folds are considered for training. For each algorithm different Performance evaluation parameters and confusion matrix are obtained.

• **Naive Bayes Algorithm**: The confusion matrix obtained after implementation of Naive Bayes classification algorithm is shown in table 3.

	Correct Labels	
	Positive	Negative
Positive	11107	1393
Negative	2834	9666

Table 3: Confusion matrix for Naive Bayes classifier

	Precision	Recall	F-Measure
Negative	0.80	0.89	0.84
Positive	0.87	0.77	0.82

Table 4: Evaluation parameters for Naive Bayes classifier

Maximum accuracy achieved after the cross validation analysis of Naive Bayes classifier is **0.895**

5.3 Analysis

Classifier	Proposed Approach
Naïve Bayes	0.895

6. CONCLUSION

In this study attempt has been made to classify sentimental reviews using machine learning techniques.

We conclude that using the different NLTK classifier it is easier to classify the reviews and more we improved the training data set more we can get accurate results. Future work should focus on training the classifier more and making the distribution proportionate to maximize the efficiency.

References

[1]. S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani, “Automatically determining attitude type and force for sentiment analysis,” in *Human Language Technology. Challenges of the Information Society*. Springer, 2009, pp. 218–231.

[2]. Y. Singh, P. K. Bhatia, and O. Sangwan, “A review of studies on machine learning techniques,” *International Journal of Computer Science and Security*, vol. 1, no. 1, pp. 70–84, 2007.

3. R. Feldman, “Techniques and applications for sentiment analysis,” *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.

[4]. B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics, 2002, pp. 79–86.

[5]. P. D. Turney, “Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,” in *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002, pp. 417–424.

[6]. K. Dave, S. Lawrence, and D. M. Pennock, “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews,” in *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003, pp. 519–528.

[7]. B. Pang and L. Lee, “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts,” in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004, p. 271.

[8]. C. Whitelaw, N. Garg, and S. Argamon, “Using appraisal groups for sentiment analysis,” in *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 2005, pp. 625–631.

[9]. S. Li, Z. Wang, G. Zhou, and S. Y. M. Lee, “Semi-supervised learning for imbalanced sentiment classification,” in *IJCAI Proceedings- International Joint Conference on Artificial Intelligence*, vol. 22, no. 3, 2011, p. 1826.

[10]. Y. Wang and X.-J. Wang, “A new approach to feature selection in text classification,” in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, vol. 6. IEEE, 2005, pp. 3814–3819.

[11]. H. Tang, S. Tan, and X. Cheng, “A survey on sentiment detection of reviews,” *Expert Systems with Applications*, vol. 36, no. 7, pp. 10 760– 10 773, 2009.

[12]. R. Garreta and G. Moncecchi, *Learning scikit-learn: Machine Learning in Python*. Packt Publishing Ltd, 2013.

[13]. A. McCallum, K. Nigam *et al.*, “A comparison of event models for naive bayes text classification,” in *AAAI-98 workshop on learning for text categorization*, vol. 752. Citeseer, 1998, pp. 41–48.

[14]. K. Mouthami, K. N. Devi, and V. M. Bhaskaran, “Sentiment analysis and classification based on textual reviews,” in *Information Communication and Embedded Systems (ICICES), 2013 International Conference on*. IEEE, 2013, pp. 271–276.

[15]. P. dataset. Polarity dataset version 2.0, sentiment anaysis dataset. [Online]. Available: http://www.cs.cornell.edu/people/pabo/ movie-review-data/review polarity.tar.gz

[16]. S. Amir, M. Almeida, B. Martins, J. Filgueiras, and M. J. Silva, "Tugas: Exploiting unlabelled data for twitter sentiment analysis," *SemEval 2014*, p. 673, 2014.

[17]. J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proceedings of the ACL Student Research Workshop*. Association for Computational Linguistics, 2005, pp. 43–48.

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

[19] Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst 2013;28:15–21.