An Improvised Feature-Based Method for Sentiment Analysis of Product Reviews

A. K. Yadav[1], D. Yadav[1] and A. Jain[2],*

[1]CSE, National Institute of Technology, Hamirpur, Himachal Pradesh, India
[2]CSE, Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India

Abstract

In today's society, sentiment analysis has gained due importance as it provides useful information about products that are used by variety of users. It gives a sneak peek of users' reactions towards the products that are available in the market at an early stage. It thus intimates users' perception and charts out a path that is beneficial for the market to grow as a whole. Although a lot of research is done to exploit the product based sentiment analysis but due to increase demand of the detailed components based products and their associated features, a novel method is desired to meet these criteria. So far, no such method is explored that analyses the product's components and their features simultaneously, on the basis of sentiments of the users. This paper proposes an improvised Feature Based Algorithm (FBA) for the sentiment analysis of product reviews while formulating a tree structure of product, components, and associated features. In addition, evaluation of double negative sentences, detecting questions and emotions from the review sentences are measured which increases efficiency of the FBA method. The comparison of product's components reviews is done with other existing algorithms-TF, TF-IDF and Naïve Bayes to demonstrate that the proposed FBA is coherent and auspicious.

---

*Corresponding author. Email: ajain.jiit@gmail.com

1. Introduction

It is quite visible to all of us that social media is growing at an explosive rate. Everything is just a click away from the users. We just think, and our thoughts turn into reality with the help of the Web. Moreover, decision making is a genre which is being thoroughly supported by the information present on the Web. The information over the internet serves a perfect guide to us and eases our decision-making abilities. This has certainly beaten up the statistics of one being dependent upon- surveys, opinion polls etc. to determine the market trends that are functioning at a particular time. The tool of internet and information available over there has given a different outlook on users' perspective for measuring dynamics of the market, in particular. There is an abundance of such information available publicly. The fact is that there is a huge reservoir of information available on the web which makes it very difficult to process and analyse according to users' needs. Therefore, there is a need to exploit more in the field of sentimental analysis so that users comfort level is maintained with ease [4].

Sentiment analysis [2] takes the backdrop of Natural Language Processing (NLP). The sentiments can be defined as ones feelings or thoughts that are expressed in words on the web, in the form of opinions. Marketing and customer services have absorbed the benefits of sentiment analysis to shape up the market scenario in the recent times. Applications of sentiment analysis have wide coverage as it determines users' behaviour and attitude towards the products that they use in their day-to-day activities [3] [15]. There is no surprise that inception and rapid growth of different fields in the sentiment analysis run in parallel with

those of social media that are used widely today on the web [11].

It is seen that while expressing opinions, people do not explicitly express them; instead they express mixed emotions in their opinions, as they talk about number of features of a product at a time. Earlier works are primarily focussed upon giving a qualitative result regarding a specific product, making it ambiguous for the users' to decide whether the product is good or bad. For example, consider an opinion- *"Sony's multimedia features are bad, but the sound quality is amazing".* This opinion has two outlooks. One is expressed in a positive way and the other in a negative way. Hence, it is important to break down the opinions according to different features and then categorize them to get the overall polarity, either positive (+ve) or negative (–ve) [1]. It is essential to do proper classification which expresses the feature specific details of a product, and is beneficial for both- seller and buyer, to determine the best product that is available in the market according to users' feedback in the form of their opinions. The users' reviews play a pivotal role in determining the potential customers and market trends for the product [20] [21]. In fact, users' behaviour is largely governed by the opinion they perceive from their surroundings. Other decisions also influence their opinions to a certain extent. It is observed that business thrives on knowing users opinion about a product in the market, so that sellers plan suitable strategy to decide cost of the product and be ready to face competition from other rival companies. Overall, dependency of one person on another person is decreased since easily available, and reliable Internet source has made everyone's life more comfortable. The internet is touching all spheres of activities and one feels a great void if it is removed from our daily activities. Easy availability of users opinion on the web has broadened the thinking of users to a much larger extend where tedious tasks are performed within a short span of time that too with ease and much lesser efforts.

In this paper, an improvised Feature Based Algorithm (FBA) for sentiment analysis of product reviews is proposed while using the polarity of the components of a product. The FBA utilizes a tree structure to define the polarity of words and of the review sentences as a whole. Experimentation on FBA depicts that the component-based product reviews outperforms the other existing techniques.

The work in this paper is organized in the following manner. Section 2 deals with the related work for the sentiment analysis. Section 3 proposes method to determine the feature-based sentiment analysis for the product's components. Section 4 shows results and performance analysis of the proposed method in comparison to other algorithms. Finally, conclusion and future work is covered in section 5.

2. Related Work

The field of sentiment analysis is exploited by several researchers in the past, for example, Shapiro et al. [8] have discussed about the news sentiments. However, focus of this research is to introduce a novel FBA method based on sentiment analysis for the product's components, in order to extract useful information from users' opinion. Before detailing the proposed FBA method, other earlier related works in the field of sentiment analysis are stated here.

Earlier, researchers have worked out different approaches for the compilation of sentiments. One of them is the manual approach which is never used alone because of its drawbacks of being very time consuming and demanding of the larger incentives. It is often chained to other practices such as workable tools. One such important tool is the Part-Of-Speech (POS) [7] tagger that helps in classifying words of sentence into their parts of speeches. The tagger classifies words into one of the eight basic POS tags- noun, pronoun, adjective, verb etc. Each word of a sentence has a syntactic role associated with it. The words once classified into their respective POS give- positive, negative or neutral sentiments that are associated with them. The polarity is defined by matching words in a sentence with the words in a dictionary. A dictionary is made to determine the polarity of words (adjectives) involved in the opinions. The dictionary contains much anticipated and commonly used adjectives along with their defined polarities. On storing them in the database, an algorithm is then used to match words with the polarities as are defined in the dictionary. The positive and negative polarity is defined accordingly, and this is done with the help of a tree structure [24]. The polarities are appended at the end of each feature of the pre-defined product tree. Consequently, a sum of all polarities is taken at the end to get to the desired result.

Two types of approaches are used for the sentiment analysis [25], namely- lexicon analysis and machine learning. Lexicon analysis is useful for smaller data set while machine learning is useful for the larger dataset. Lexicon analysis uses polarity of text based on lexicons, such as Word Sense Disambiguation (WSD) based SentiWordNet [23]. Machine learning is classified as supervised and unsupervised. In the supervised machine learning sentiment analysis, data is trained on some standard classification, while in the unsupervised machine learning sentiment analysis no training is done on the tagged data. In the unsupervised mode, although output is not predicted but identification of sentiments from the blog contents is done using emotions like "sorrow", "happy", "anger" and "fear". It helps search engine to find statistics on the basis of these four emotions.

Hassan et al. [6] have illustrated the classification of tweets while labelling them- positive, negative or neutral according to the sentiments or opinions expressed by them. The classification is done based on the fact that if something good is projected by a tweet then it is termed as positive, else negative. Today, companies are largely dependent upon this feature as they manufacture products according to the market trends. Thus, the analysis of sentiments carries a huge economic aspect along with it. The users' reactions prompt many companies to decide their strategy based upon it. Jain et al. [22] predict the prices of Bitcoin and Litecoin using sentiments from the users' posted tweets related to the two types of cryptocurrencies. The Multiple Linear

Regression (MLR) determines the price of the cryptocurriencies with R2_score of 44% for Bitcoin and 59% for Litecoin respectively. Ahmad and Laroche [13] have grouped the sentiment analysis into few categories based on the occurrence of discrete emotional words such as "hope", "happiness", "anxiety" and "disgust" while adopting the Latent Semantic Analysis (LSA) as a quantitative content analysis approach. Gupte et al. [10] have compared different classification algorithms for the sentiment analysis such as Naïve Bayes, Maximum Entropy, Boosted Trees and Random Forest. They have concluded that whenever accuracy is the selection criteria then Random Forest is the chosen, however, it has higher learning time. On the other hand, whenever processing power and memory are at the highest priority then Naïve Bayes is selected. Whenever lesser training time is the highest priority with the powerful processing and memory then Maximum Entropy is proven to be worthy. Otherwise, there is a need to select an algorithm that is average on all aspects, and then Boosted Trees is the right choice. The selection of an appropriate classification algorithm in sentiment analysis is must as its decision influences the behavior and end-product of the system. Chakraverty and Saraswat [14] have proposed Cross Domain Recommender (CDR) system which linkages source and target domains using users' emotional categories. These categories are- "love", "joy", "anger", "surprise", "sadness" and "fear". They have experimentally verified that CDR based sentiments from users' reviews over Movielens and Bookcrossing datasets provide 77.1% better F-Score. Panthati et al. [19] have proposed a method for the sentiment analysis of products using deep learning. The method combines word2vec and Convolutional Neural Network (CNN) to predict the review sentiments from Amazon for the mobile phone product. The word2vec model converts text to word embeddings. The CNN model implements Keras and Tensorflow as deep learning frameworks. The method has better accuracy which increases substantially on increasing the size of the dataset.

The coming section highlights the proposed tree structure and method that is used to define the polarity of words and of the sentences as a whole. This is done to enable the improvised FBA based sentiment analysis of product reviews, using polarity of features components of the product.

3. Proposed Feature Based Algorithm

To access data, users' opinions are retrieved from the related product websites. The selection is product-wise. After retrieving, it may be a group of result or some unpredictable results. The opinions are processed on an individual basis to obtain the polarity of different components of the product. On obtaining an opinion, it is assessed on sentence basis. The conjunctions and full-stops are used to indicate the end of a particular sentence. One sentence is taken at a time and each word of it is targeted individually which are classified into POS categories using exiting tools- POS tagger [18] and Named Entity Recognition (NER) [5] [16].

In this work, perseverance is towards the improvised Feature Based Algorithm (FBA) for the components of a product which are in a given review from the user. This gives us a categorization of different parts of speech. The individual nouns that are obtained with the help of the said tools may contain some unwanted information as well, like names of person, place, animal etc., which are of no use. Such unintentional information has no contribution towards the components of the product, and so is considered useless. Therefore, to obtain the nouns containing the product components and their features, for example, consider a product "Mobile" whose components are "Battery", "Keypad", "Screen" etc. and the associated features are- "Memory Size", "Durability" etc. which are matched with the data. The data is matched on the defined data structure of the product and index of the features, regarding which the review sentence is written by the user is taken into account. In case, the review sentence is giving information about more than one feature of the product then the index number of all the features are taken in the queue.

The data structure that is applied for the proposed FBA method is a "Tree". For example, Figure 1 depicts the tree data structure whose root node is taken as "Mobile". The root has several child nodes such as "Battery", "Screen", "Camera", "Processor", "Speaker" etc. These children nodes have subsequent attributes/features of these components as their further child nodes. These children (features) of the previous child nodes (components of the product) contain adjectives POS tag such as "Good", "Bad" etc. The leaves of the tree represent polarity of the adjectives. For example, "Good" has value "+1" while "Bad" has value "-1". This proposed tree structure provides an aerial view of all the levels (e.g. root-child-child-leaf nodes) that are involved to get the polarity. In the end, the feature-specific polarity is computed as the probability measure which is displayed as the result.
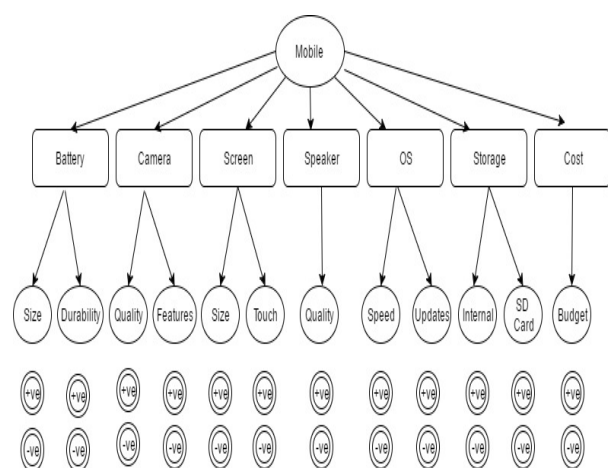

Figure 1. Tree with subsequent positive and negative polarities of the features of the product components

Figure 1 depicts the components of the product with their features and along with their subsequent polarities. The values of these polarities are obtained with the help of the dictionary, which stores the parts-of-speech words and their assigned numbers (polarities). It is now seen that when all the sentiments are processed, the polarities of each feature are assembled below it. Now, the count is made easily. There arise some cases e.g. the quality of camera gets a negative polarity while features of the camera get a positive polarity. Here, the classification is made much easier by assigning such polarities to the numbers. Another example, screen size is favourable but the touch screen is not favourable at all. The entire review is taken initially and is segregated into sentences, which are further segregated into different words with the help of different tools. To aid the polarity building strategy, help of the dictionary is taken into consideration. The sample dictionary words along with their polarity that are used in the FBA are shown in Table 1.

Table 1. A sample dictionary words along with polarity.

| SNo. | Dictionary | |
	Words	Polarity
1	Good	+1
2	Better	+2
3	Best	+3
4	Bad	-1
5	Worse	-2
6	Worst	-3
7	Exceptional	+4
8	Deplorable	-4
9	Amazing	+5
10	Pathetic	-5
11	Outstanding	+6
12	Dreadful	-6

For example in a given sentence: "The screen size is worst but the battery is outstanding". Now the nouns that are obtained here are: "screen" and "battery". The indices of these product components are obtained from the pre-defined tree structure. Then an index is selected at a time, and its corresponding feature is targeted so that the adjective following the noun in the sentence are picked up. The proposed adjectives are then matched with the words in the dictionary to obtain their polarity. The value of polarity is appended into nodes which are duly attached below the node, corresponding to the feature in the tree. Therefore, on attaining their polarity from the dictionary, their value is added to the corresponding node in the tree.

3.1 Pseudo code and Flowchart

Assume that all reviews are as input file and create dictionary of review and store it as tree structure.

Step 1: Store all reviews in the form of string.

Step 2: Differentiate all part of speech words.

Step 3: Add synonyms of positive and negative words in dictionary.

Step 4: Add questionable keywords and its symbol.

Step 5: Assign polarity to the adjectives.

Step 6: Find the node of tree and update the tree.

Step 7: Generate the results.

Figure 2 shows the flowchart of the proposed FBA in which steps are repeated for all the sentences in the opinion obtained from the user. Similarly, all the opinions are processed. The FBA tree gets updated with the values of the negative and positive polarities of all the features mentioned. Next, the cumulative result is obtained. The task is performed for various opinions and at the end, after appending the polarity of each adjective to its corresponding node in the tree, a sum of all the polarities is taken and the final polarity value of each feature of the product is obtained.

As discussed in the Algorithm, the tree structure first finds the nouns from reviews followed by adjectives of related nouns. Retrieved adjectives are compared with dictionary list and assign weight to each adjective. In the last step, sentiment classification followed by sentiment evaluation have been done.
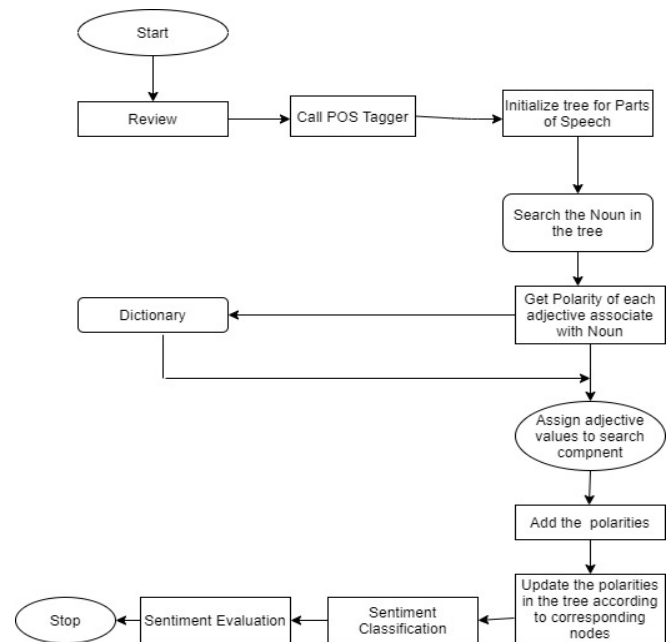

Figure 2. Flowchart of improvised FBA

Example: Let us consider 7 reviews from a website for the product component- "Operating System (OS)" for the

product- "Mobile". It has found that there are 4 positive and 3 negatives reviews. The proposed feature based sentiment analysis algorithm is applied to find the polarity of the product, as follows.

Total reviews for the product OS = no. of positive + negative reviews = 4 + 3= 7

Probability (OS being liked, overall) = No. of positive reviews / total reviews = 4/7

Probability (OS being disliked, overall) = No. of negative reviews / total Reviews = 3/7

In this way, calculate the probability of being liked and disliked for each component and this gives the cumulative polarity of the components of the root "Mobile".

4. Result Analysis

Consider the sentiments of the users, the reviews of their products are taken from commercial websites. In the present case, reviews taken are retrieved from Flipkart [9], Amazon [27] and Tata CLiQ [28] in the month of November and December 2019. The detail of source reviews, numbers of positive and negative reviews are shown in Table 2.

Table 2. Source Reviews

Data Source	No. of Reviews		Total
	Positive	Negative	
Flipkart	1300	577	1877
Amazon	1000	410	1410
Tata CLiQ	1100	315	1415

By using this dataset, we have divided the data set into two parts (1) Training Set (2) Testing Set.

Training Set: The training set contains 3796 reviews that are divided into six aspects as shown in Table.3.

Table 3. Training Set

	Positive	Negative	Total
Mobile	935	267	1202
Camera	185	129	314
Battery Back-up	338	251	589
Screen	284	143	427
Battery-Heat	864	222	1086
Operating System	160	18	178

Test Set: The test set contains 906 reviews with six aspects as shown in Table 4.

Table 4. Test Set

	Positive	Negative	Total
Mobile	98	10	108
Camera	110	14	124
Battery Back-up	116	14	130
Screen	128	30	158
Battery-Heat	144	52	196
Operating System	150	40	190

As per discussion in Algorithm 1, the prediction of each noun with features of adjectives. It may be positive or negative. The prediction is calculated by comparing the adjectives with the dictionary word and set the polarity according to values assigned in dictionary (Table.1). The final outcome will be sum of all the polarity (positive and negative). For thresholding, its class is defined as positive if its sum is +2 or more, otherwise it is assigned as negative.

During the training phase, the rules have been identified for sentiment orientation, considering conjuncts, to improve the efficiency of tweets. Different reviews are considered and scores are given to each feature by sentiment evaluation. During evaluation, TF, TF-IDF value is calculated for each sentiment word present in review. Using dictionary system (as shown in table Table.1), the sentiment words are given weights and calculate the overall score of reviews. If the overall score is +2 or more than it is considered as class positive otherwise negative class.

Performance of the proposed FBA algorithm is evaluated using precision, recall and F-Score for reviews of the product- "Mobile", its components, their associated features respectively.

Precision is defined as the number of correct predictions. It is a ratio between the number of correct results displayed and number of total results displayed, as in equation (1).

$$Precision = \frac{Correct\ Predictions}{Total\ Number\ of\ Results} \quad (1)$$

Recall is defined as how many relevant items are selected. It is a ratio of the number of correct results displayed and total number of correct results that are available in the system, as in equation (2).

$$Recall = \frac{Correct\ Predictions}{Total\ Number\ of\ Correct\ Results} \quad (2)$$

F-Score (F) is defined as the harmonic mean between the precision and recall, as in equation (3).

$$F = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

TF is the frequency of occurrence, of a word or group of words in a review. This is also called Bag of Words model. In this model, each document is represented as a vector of 0s and 1s. If a word exists in a document, its corresponding position in the vector is coded as a "1" and if it doesn't, it is coded as a 0.

TF is calculated as follow:

$$TF = \frac{Frequency\ of\ words\ in\ the\ review}{Number\ of\ word\ in\ the\ review} \quad (4)$$

The IDF of a word is the measure of the relative importance of that word is in the whole corpus.

IDF is calculated as follow:

$$IDF(word) = \log \frac{Total\ no.of\ reviews}{Number\ of\ reviews\ containing\ the\ word} \quad (5)$$

TF-IDF is the product of TF and IDF score for specific words. In TF-IDF model, each document is represented as a vector that contains TF-IDF scores for each of the words in the review. TF-IDF scales down the impact of frequent but less informative features.

$$TF - IDF = TF * IDF \quad (6)$$

The precision and recall measures are calculated on the improvised FBA and are compared with TF [12], TF-IDF [17] and Naïve Bayes [26] algorithms.

Table 5. Precision of FBA

Mobile Dataset	Total Reviews Retrieved	Sentiment Polarity Positive	Sentiment Polarity Negative	Precision (Positive)
Mobile	108	98	10	0.91
Camera	124	110	14	0.89
Battery Backup	130	116	14	0.89
Screen	158	128	30	0.81
Battery Heat	196	144	52	0.73
Operating System	190	150	40	0.79

Table 5 considers the Mobile dataset where total reviews retrieved, sentiment polarity (positive and negative) are given. It then computes the precision measure of positive sentiments for different components of Mobile such as "Camera", "Battery Backup", "Screen", "Battery Heat" and "Operating System" respectively.

Table 6. Recall of FBA

Mobile Dataset	Total Reviews (Positive)	Relevant Document (Positive)	Recall (Positive)
Mobile	107	54	0.51
Camera	112	62	0.55
Battery Backup	121	65	0.54
Screen	133	79	0.59
Battery Heat	147	98	0.67
Operating System	143	95	0.66

Table 6 considers the Mobile dataset where total positive reviews retrieved, relevant positive documents are given. It then computes the recall measure of positive sentiments for different components of the Mobile.

Table 7. Performance evaluation of FBA with TF, TF-IDF and Naïve Bayes using precision and recall

Mobile Dataset	Method	Positive Sentiments Precision	Positive Sentiments Recall	Negative Sentiments Precision	Negative Sentiments Recall
Mobile	TF	0.82	0.91	0.88	0.80
	TF-IDF	0.86	0.93	0.82	0.83
	Naïve Bayes	0.89	0.88	0.86	0.81
	FBA	0.91	0.51	0.88	0.48
Camera	TF	0.78	0.69	0.76	0.47
	TF-IDF	0.81	0.81	0.71	0.48
	Naïve Bayes	0.87	0.79	0.76	0.59
	FBA	0.89	0.55	0.86	0.54
Battery Backup	TF	0.82	0.59	0.81	0.57
	TF-IDF	0.76	0.61	0.78	0.52
	Naïve Bayes	0.81	0.52	0.63	0.61
	FBA	0.89	0.54	0.91	0.51
Screen	TF	0.63	0.51	0.51	0.61
	TF-IDF	0.68	0.48	0.65	0.51
	Naïve Bayes	0.79	0.51	0.81	0.49
	FBA	0.81	0.59	0.85	0.62
Battery Heat	TF	0.73	0.52	0.81	0.63
	TF-IDF	0.82	0.63	0.76	0.53
	Naïve Bayes	0.71	0.53	0.71	0.57
	FBA	0.73	0.67	0.71	0.62
Operating System	TF	0.73	0.59	0.69	0.58
	TF-IDF	0.74	0.60	0.70	0.59
	Naïve Bayes	0.76	0.61	0.69	0.57
	FBA	0.79	0.66	0.71	0.60

Table 7 compares precision and recall measures of FBA with TF, TF-IDF and Naïve Bayes for positive and negative polarities over the Mobile dataset. It is seen that for most of the components- "Camera", "Battery Backup", "Screen" and "Operating System" the positive sentiment of precision is quite higher for FBA while for the "Battery Heat" it is equivalent to TF. The recall of positive sentiment for "Screen", "Battery Size" and "Operating System" are higher for FBA while lower for "Camera" and "Battery Backup". On the other hand, precision for negative sentiment are higher for FBA, for all components except for "Battery Heat". The recall of negative sentiment for "Screen", "Battery Heat" and "Operating System" are higher for FBA while it is lesser than TF-IDF mainly for other components.

Table 8. Performance evaluation of FBA with TF, TF-IDF and Naïve Bayes using F-Score

Mobile Dataset	Method	F-Score	
		Positive	Negative
Mobile	TF	0.83	0.84
	TF-IDF	0.90	0.82
	Naïve Bayes	1.12	0.83
	FBA	0.61	0.62
Camera	TF	0.66	0.58
	TF-IDF	0.79	0.57
	Naïve Bayes	0.96	0.66
	FBA	0.69	0.65
Battery Backup	TF	0.71	0.67
	TF-IDF	0.70	0.62
	Naïve Bayes	0.59	0.62
	FBA	0.83	0.65
Screen	TF	0.55	0.56
	TF-IDF	0.50	0.57
	Naïve Bayes	0.58	0.61
	FBA	0.76	0.72
Battery Heat	TF	0.52	0.71
	TF-IDF	0.83	0.62
	Naïve Bayes	0.54	0.63
	FBA	0.73	0.66
Operating System	TF	0.64	0.63
	TF-IDF	0.65	0.64
	Naïve Bayes	0.65	0.62
	FBA	0.83	0.65

Table 8 compares F-Score measures of FBA with other algorithms- TF, TF-IDF and Naïve Bayes for both the positive and negative polarities over the Mobile dataset. F-Score of positive sentiments for "Battery Backup", "Screen" and "Operating System" are quite higher for FBA over TF,

TF-IDF and Naïve Bayes. However, F-Score of negative sentiments for "Screen" and "Operating System" are quite higher for FBA in comparison to other algorithms. Overall, evaluation has shown that the FBA clearly outperforms and provides better results in comparison to the other existing algorithms.

5. Conclusion and Future Work

In this paper, it is observed that sentimental analysis plays a vital role in decision making for the users about the choice of a product. To do so, an improvised Feature Based Algorithm (FBA) is proposed which uses a tree data structure while looking insight of a product, its components and associated features. For each product, indices of the product components are obtained from the pre-defined tree structure. An index is selected at a time, and its corresponding features are targeted, so that features (adjectives POS) following the product components (noun POS) in the review sentences are picked up. The adjectives are then matched with the words in the built-in dictionary to obtain their polarity- positive or negative. The FBA tree gets updated with the polarity values of features mentioned in the review sentences. Next, the cumulative result is obtained. The task is performed for various sentiments, and at the end, after appending the polarity of each adjective to its corresponding node in the tree, a sum of all the polarities is taken. In this research, 4702 reviews are retrieved from the Flipkart, Amazon and Tata CLiQ websites during 2019. The performance of the FBA is compared with TF, TF-IDF and Naïve Bayes using precision, recall and F-Score for both the positive and negative sentiments. The performance of FBA is significantly better than other algorithms which benefits users for choosing the products.

In the future, sentimental analysis for product reviews can be explored using the amalgamation of FBA with the data mining rules. Some more issues like identification of fake reviews and their elimination can also be explored.

References

[1] G. Gautam, D. Yadav. Sentiment analysis of Twitter data using machine learning approaches and semantic analysis. In Proceedings of the 7th International Conference on Contemporary Computing (IC3); August 2014; Noida, India. IEEE; 2014. pp. 437-442.

[2] F. Xing, J. Zhan. Sentiment analysis using product review data. Journal of Big Data. 2015; 2(5), pp.1-14.

[3] G. Khanvilkar, D. Vora. Sentiment analysis for product recommendation using random forest. International Journal of Engineering &Technology. 2018; 7(3.3), pp.87-89.

[4] M. Cheng, X. Jin. What do Airbnb users care about? An analysis of online review comments. International Journal of Hospitality Management. 2019; 76, pp.58-70.

[5] A. Jain. Named entity recognition for Hindi language using NLP techniques. 2019; Ph.D. Thesis, Jaypee Institute of Information Technology, Noida, India.

[6] S. U. Hassan, N. R. Aljohani, N. Idrees, R. Sarwar, R. Nawaz, E. Martínez-Cámara, S. Ventura, F. Herrera.

Predicting literature's early impact with sentiment analysis in Twitter. Knowledge-Based Systems. Elsevier. 2020; 192:105383.

[7] J. P. Gupta, D. K. Tayal, A. Gupta. A TENGRAM method based part-of-speech tagging of multi-category words in Hindi language. Expert Systems with Applications. Elsevier. 2011; 38(12), pp.15084-15093.

[8] A. H. Shapiro, M. Sudhof, D. Measuring news sentiment. Federal Reserve Bank of San Francisco Working Paper 2017-01. 2020. https://doi.org/10.24148/wp2017-01

[9] Flipkart Reviews. Accessed on 10-30 November 2019. www.flipkart.com

[10] A. Gupte, S. Joshi, P. Gadgul, A. Kadam. Comparative study of classification algorithms used in sentiment analysis. International Journal of Computer Science and Information Technologies. 2014; 5(5), pp. 6261-6264.

[11] A. Jain, A. Gupta, N. Sharma, S. Joshi, D. Yadav. Mining application on analysing users' interests from Twitter. In Proceedings of the 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT); March 26-27; MNIT Jaipur, India. ELSEVIER-SSRN; 2018. pp. 20-27.

[12] B. S. Rintyarna, R. Sarno, C. Fatichah. Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context. International Journal of Information and Decision Sciences. 2020; 12(1), pp.75-101.

[13] S. N. Ahmad, M. Laroche. How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. International Journal of Electronic Commerce. 2015; 20(1), pp.76-111.

[14] S. Chakraverty, M. Saraswat. Review based emotion profiles for cross domain recommendation. Multimedia Tools and Applications. 2017; 76(24), pp. 25827-25850.

[15] Q. Zhong, S. Liang, L. Cui, H. K. Chan, Y. Qiu. Using Online Reviews to Explore Consumer Purchasing Behaviour in Different Cultural Settings. Kybernetes. 2019; 48(6), pp. 1242-1263.

[16] A. Jain, D. K. Tayal, D. Yadav, A. Arora. In Data Visualization and Knowledge Engineering. Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 32). Springer; 2020. Chapter 10, Research trends for named entity recognition in Hindi language; pp. 223-248.

[17] S. Uppal, A. Jain, A. Arora. Comparative analysis for keyterms extraction methods for personalized search engines. In Proceedings of the 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence); January 2020; Noida, India; IEEE; 2020. pp. 358-363.

[18] A. Jain, A. Arora. Named entity recognition in Hindi using hyperspace analogue to language and conditional random field. Pertanika Journal of Science & Technology. 2018; 26(4): pp.1801-1822.

[19] J. Panthati, J. Bhaskar, T. K. Ranga, M. R. Challa, M. Sentiment analysis of product reviews using deep learning. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI); IEEE; 2018, pp. 2408-2414.

[20] M. Chauhan, D. Yadav. Sentimental analysis of product based reviews using machine learning approaches. Journal of Network Communications and Emerging Technologies (JNCET). 2015; 5(2), pp.19-25.

[21] S. M. H. Chowdhury, S. Abujar, K. B. M. Badruzzaman, S. A. Hossain. In Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems (LNNS, volume 103). Springer; 2020. Product review analysis using social media data based on sentiment analysis; pp. 527-533.

[22] A. Jain, S. Tripathi, H. DharDwivedi, P. Saxena. Forecasting price of cryptocurrencies using tweets sentiment analysis. In Proceedings of the 11th International Conference on Contemporary Computing (IC3); August 2018; Noida, Indai. IEEE; 2018. pp. 1-7.

[23] C. Hung, S. J. Chen. Word sense disambiguation based sentiment lexicons for sentiment classification. Knowledge-Based Systems. Elsevier. 2016; 110, pp. 224-232.

[24] S. Kim, J. Zhang, Z. Chen, A. Oh, S. Liu. A hierarchical aspect-sentiment model for online reviews. In Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence. 2013; pp. 526-533.

[25] N. Nigam, D. Yadav. Lexicon based approach for sentiment analysis of tweets using R language. In Proceedings of the International Conference on Advances in Computing and Data Sciences (ICACDS); April 20-21, 2018; Springer; 2018; pp. 154-164.

[26] A. Jain, R. Gairola, S. Jain, A. Arora. In Social Network Analytics for Contemporary Business Organizations. IGI Global; 2018. Chapter 4, Thwarting Spam on Facebook: Identifying Spam Posts Using Machine Learning Techniques; pp. 51-70.

[27] Amazon Reviews. Accessed on 10-30 November 2019. www.amazon.com

[28] Tata click Reviews. Accessed on 10-30 November 2019. www.tatacliq.com