IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. V. (Jul.-Aug. 2017), PP 74-83
www.iosrjournals.org

Extracting Topical Relation between Opinion Target and Opinion Word using Modified Naïve Bayes Approach

*Prajkta Akre[1], Harshali Patil[2], Mohammad Atique[3]
[1](Computer, Thakur College of Engineering and Technology/Mumbai University, India)
[2](Computer, Thakur College of Engineering and Technology /Mumbai University, India)
Corresponding Author: *Prajkta Akre*

Abstract: *Opinion mining involves building a system to collect reviews from online social sites and categorize opinions about a product, service. In opinion mining, finding opinion targets and opinion words from online reviews are important tasks. Key component of opinion mining involves detecting opinion relations among words. The primary purpose of this project is to make these extraction processes more effective. Additional types of relation between words, such as "Topical Relations" are considered. Topical relation denotes how well a retrieved document or set of documents meets the information given by the user. These topical relations are useful for the user to gain the insight of topics that are mostly discussed by customers for a particular product. The system also calculated sentiment of customer reviews which are classified into three categories positive, negative and neutral. To classify the reviews the modified naïve bayes classifier is used. Modified naïve bayes produces better results than the traditional naïve bayes classifier and Decision tree. For experimental results, these algorithms are compared using different performance parameters for different number of records. On the basis of the results of this research, it can be concluded that modified naïve bayes shown better precision and accuracy than the naïve bayes algorithm and decision tree.*

Keywords: *Modified naïve bayes, Opinion mining, Opinion Target, Opinion word, Topical relation.*

Date of Submission: 28-07-2017 Date of acceptance: 14-08-2017

I. Introduction

Opinion mining of online reviews about the product is very useful in today's world. Most of the customers rely on the reviews submitted by the other users to make the decision of buying the product or not. These reviews are also helpful for the companies to make their products better than the other products. Product reviews are easily available on any website, blogs and forums. But reading those hundreds of reviews requires lots of time. Sometimes customers won't get the exact information that they need. These are the some drawbacks form the existing systems. To overcome these drawbacks we provide feature based classification of the product reviews. Mostly customers look for the better features of any product before buying it. Our work will provide sentiment about the reviews of the product and also classifies those reviews based on their features.

Opinion mining or sentimental analysis requires several tasks to be performed to calculate the sentiment of the reviews. Our system performs several tasks consists of gathering reviews from the online site, pre-processing the data to process it further, extracting the opinion target and opinion words, calculating sentimental value and finally classify the reviews into positive, negative and neutral based on their features.

There are many classifier algorithms are used to classify the sentiment of the reviews. In our work, we used traditional naive bayes algorithm which is used to predict the sentiment of the review. Naive bayes classify the reviews based on the probability calculated from the trained data set.

Domain Relevance or Topical relations are the topics or subjects which are mostly discussed by the customers while expressing a review about product or a service. As we are concentrating on electronic products, topics which are mostly dealing with are battery, price, performance, memory space, platform and looks etc. These are nothing but the features of the electronic products which are discussed by the customers.

1.1 Objectives

This work will look to meet some of the following objectives:

- To design enhanced opinion mining system to find out the topics which are mostly discussed by the different users to express their opinion about particular item, product or services.
- To design the algorithm that will show the relation between opinion target and opinion word.
- To design the algorithm that will display the sentiment value of online reviews.
- To classify the reviews into their correct classes using modified naïve bayes classifier.
- To compare modified naïve bayes classifier with the simple naïve bayes.

II. Literature Review

A substantial work has been done over the past years in the area of Opinion mining.

2.1 Sentiment Analysis and Classification for Software As A Service Reviews

Asma Alkabani, Ahmed Ghamry, Farookh Hussain and Omkar Husain [1] done their work to extract reviews from the customers those are taking benefit of online services provided from cloud computing. After using those services feedback taken from users in subjective and objective manner. They used SVM model to predict the sentiments of opinions taken from users.

2.2 Opinion mining from student feedback data using supervised learning algorithms

Dhanlakshmi V, Sarvanan A. [2] implemented opinion mining using supervised algorithms to find out the sentiment about student's feedback, which is based on teaching and learning module. They have also used Rapid Miner tool to for step by step explanation of opinion mining. They did comparative study of the algorithms like SVM, Naive Bayes and Neural Network classifier. Their result shows that Naive Bayes algorithm performed very well. In their future work they wanted to gather data of a student from social media and analyse the opinion features.

2.3 Co-Extracting Opinion Targets and Opinion words from online reviews

Kang Liu, Liheng Xu and Jun Zhao [3] extracting opinion targets and opinion words based on word alignment model which is a partially supervised alignment model. They also used graph-based co-ranking algorithm to calculate the confidence of candidate features, candidate features which are having higher confidence are considered as the final opinion targets and opinion words for the extraction. To show the effectiveness of their proposed method, they used three datasets with different languages with different data sizes.

2.4 Sentiment Analysis for Social Media: A Survey

Harshali Patil, Mohammad Atique [4] made survey of various research papers which shows challenges in opinion mining which are faced by most of the researchers. Main challenge is to analyse the real time data. The data set which requires data that changes over time must use incremental approach, which allows past updating of past results using new data instances. Another challenge is to tackle grammatical errors. We can improve our results by mapping theses grammatical errors to correct words. As we are dealing with online reviews, there are chances that the data can be noisy, unstructured and dynamic in nature. Removal of such noisy data is also a challenging task.

2.5 Extracting Topical Information of Tweets Using Hash tags

Zeynep Alp and sule [5] used twitter data to perform opinion mining and to extract topical information of various tweets submitted by different users. Unlike any research they did not collect tweets which are based on queries and keywords. They collected tweets from group of people following friend and follower information starting from a specific user. They developed proposed method to overcome the challenge of tweets being short for topical modelling. They used various pooling schemes, to extract the topical information from tweets. From their experiments they had shown that pooling tweets together improves LDA performance with unbiased data collection and evaluation methodologies.

2.6 Drawbacks of the Existing Systems

- The methods applied were effective for opinion target extraction but those methods were unsatisfactory for opinion word extraction.
- Existing methods did not worked on "Topical Relation" between opinion target and opinion words.

III. Proposed System

The proposed architecture for the Opinion Mining System is shown below.

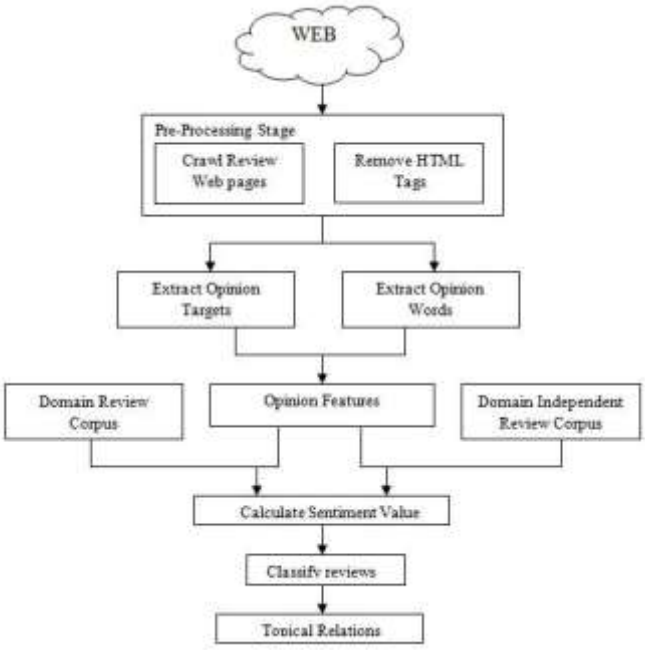

Fig. 3.1 Architecture of Proposed System

3.1 Module 1: Extraction of reviews from web pages

This module extracts thousands of electronic product reviews from online sites such as Amazon.com and Flipkart.com.

3.2 Module 2: Pre-processing stage

This module makes our system effective by removing stop-words and using stemming process. Stop-words are common words which are frequently used in most of the reviews such as 'is', 'the' and 'a' etc. Stemming process will reduce derived words to their base form.

3.3 Module 3: Extraction of Opinion Target and Words

Consider a review "Samsung galaxy j2 battery is really good….display is also excellent". In this review, Samsung galaxy j2 represented opinion target and adjectives such as good, excellent are opinion words which represent positive impact on the product. Different dataset created to store opinion target and opinion words.

3.4 Module 4: Calculating the confidence

To calculate the confidence, we consider three types of opinion words. Opinion words can be expressed using Adjective, Adverb and Verb. We calculated no. of adjectives, adverbs and verbs present in each statement.

adjcnt= count of adjective per sentence……(1)
advcnt= count of adverb per sentence……(2)
vbcnt= count of verbs per sentence…….(3)

3.5 Module 5: Extraction of Opinion features

To shown relation between opinion target and opinion words, it is important to extract valid opinion features. Proposed system works on electronic reviews, basic opinion features that are extracted are battery, looks, display, performance etc. To extract such opinion features, system follows set of rules which are listed below.

Table 1. Rules for Extraction of Opinion Features

Rules	Interpretation
NN + SBV = CF	Identify NN as a CF,If NN has a SBV Dependency relation
NN + VOB = CF	Identify NN as a CF,If NN has a VOB Dependency relation
NN + POB = CF	Identify NN as a CF,If NN has a POB Dependency relation

Consider the same example, "It has a nice colour screen". In this review nice is adjective which is associated with screen colour, colour is noun which has subject verb dependency on "screen". So here, screen

will be extracted as a feature of the product. Using these rules we are able to extract remaining features from online review document.

3.6 Module 6: Calculating Sentiment value

For calculating sentiment, we assign value to adjectives, adverbs and verbs according to their polarity such as good=0.4 worst=-0.8 very=0.65 recommend=0.8. After assigning values to the words, positions of these words are found out to check whether there is any co-relation exists between them. Proposed system consider two types of co-relation and those are adjective –adverb co-relation and adjective-verb co-relation.

3.7 Module 7: Classification of reviews

For classification of electronic reviews, naïve bayes and modified naïve bayes algorithm used.

3.7.1 Classification using Naïve Bayes algorithm

After calculating sentiment, we checked the probabilities of positive, negative or neutral reviews. If probability of positive count is greater than the negative count then we classify our review into positive class or vice versa.

For feature based classification, we calculated probability of every sentence against the extracted features. If the probability of particular sentence for battery feature is greater than the probability of display, then we classify our sentence into battery feature. If the particular sentence contains two features for example, "Display is excellent...it also has a good battery....I love this phone." In this sentence there are two features of the phone display and battery, so naive bayes classifier will classify this sentence into both features. This is the drawback of the naive bayes, which is removed using modified naive bayes.

3.7.2 Classification using Modified Naïve Bayes algorithm

Modified naïve bayes algorithm, will classify the reviews according to their opinion weightage. Consider above example "Display is excellent….it also has a good battery….I love this phone." In this review, display feature having more positive value than the battery feature. Modified naïve bayes algorithm, compare both the values given to features, and feature with higher value classifies in its class. Above review will classify in display feature rather than battery feature.

IV. Experimental Setup

4.1 Dataset Description

The proposed system two types of datasets, one is training dataset which contains thousands of reviews about electronic product and another is test dataset which contains reviews according to electronic products. These reviews are collected from online sites such as www.flipkart.com and we also take dataset from one of the author who extracted product reviews from www.amazon.com. Reviews that we collected may contain short comments, complete sentence or paragraph.

The system provides user friendly interface for adding, updating or deleting user review information.

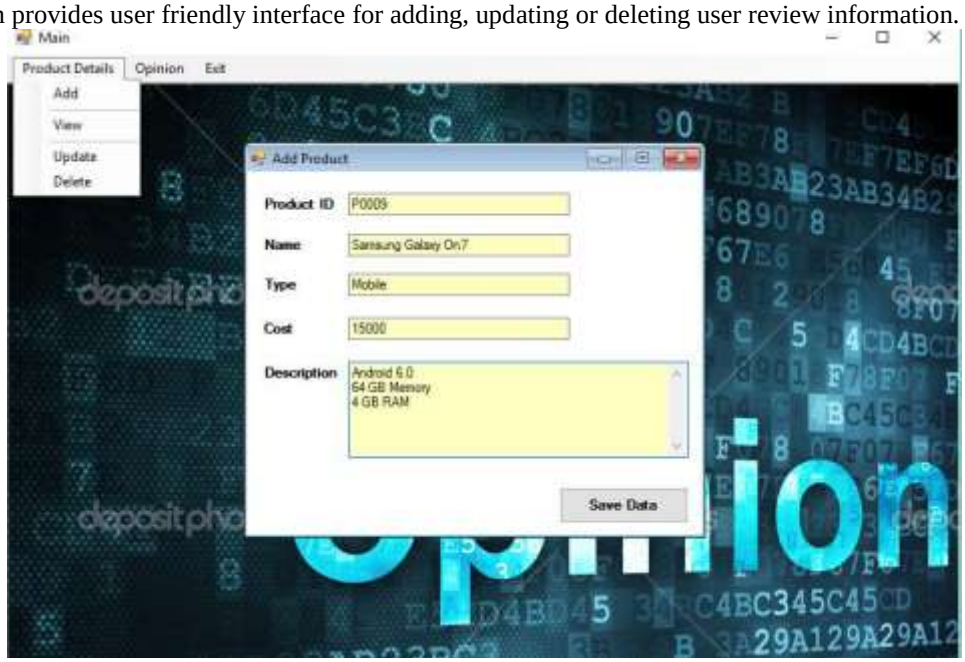

Fig. 4.1 Form for adding product details

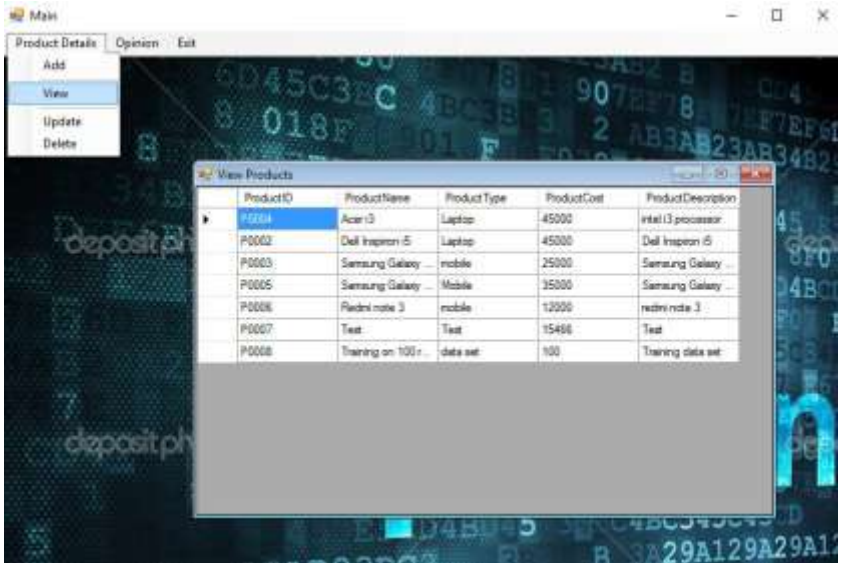

Fig. 4.2 Form to view product details

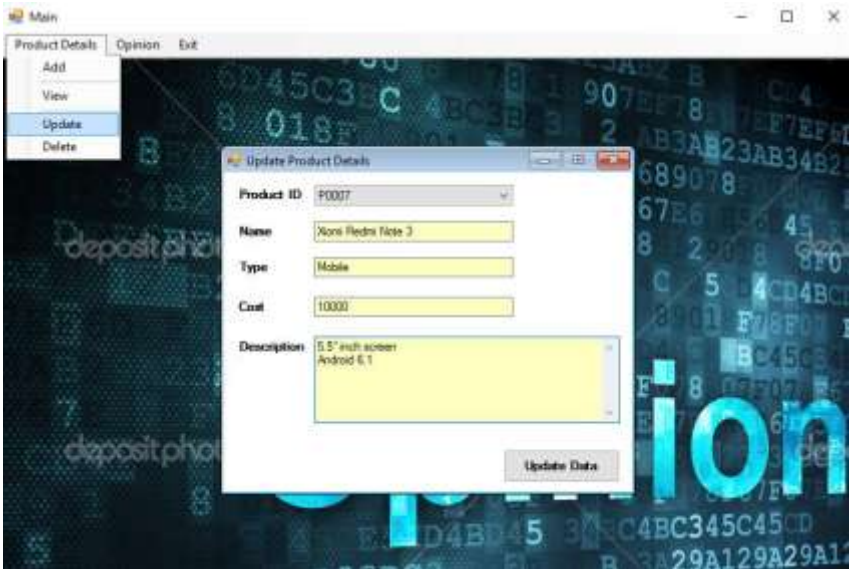

Fig. 4.3 Form to update product details

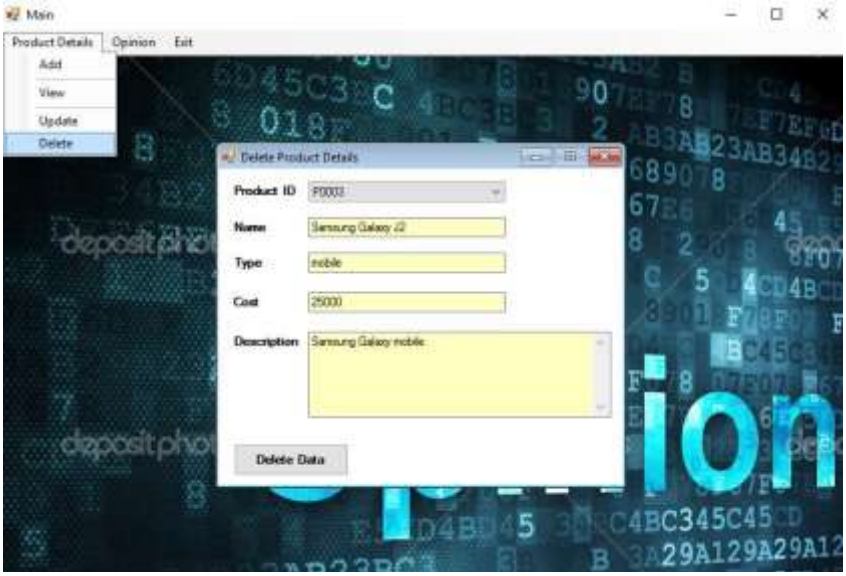

Fig. 4.4 Form to delete product details

4.2 Results obtained by proposed system

After adding products, now we will add reviews of that product. To add the product reviews, click on Add menu, under that select NB menu i.e. Naive Bayes.

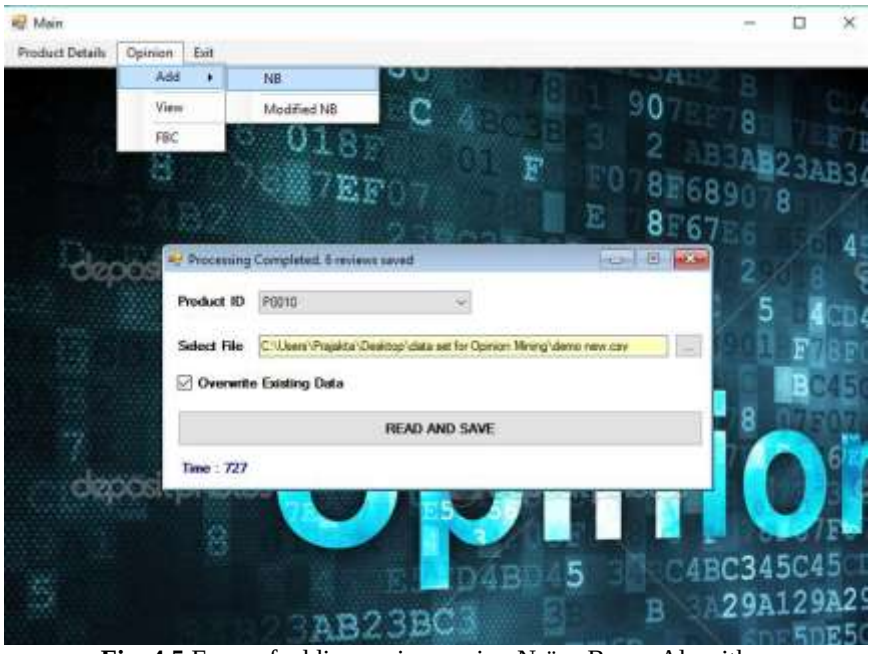

Fig. 4.5 Form of adding reviews using Naïve Bayes Algorithm

As shown in above fig. 4.2, after clicking read and save button system shows execution time which includes process of reading reviews and calculating sentiment of those reviews. To see the result of naïve bayes algorithm, select view menu.

Fig. 4.6 Result of reviews using Naïve Bayes Algorithm

As shown in above fig. 4.6, view form contains three columns, First column shows reviews and sentiment values of all products which are already loaded in the system. To see the specific product reviews, select product id in second column. We select product id P0010 from the drop down list and we get all the reviews and their sentiment value related to P0010. Third column contains different features of the product such as battery, looks, display etc. When we select any of the features from the list, we get review of product P0010 belonging to that feature. For example, we select "Battery" feature in above example then we get review related to battery feature such as "As usual battery power is not sufficient for daily usage" with its sentiment value i.e. 0 (Neutral).

Fig. 4.7 Result of reviews using Naïve Bayes Algorithm

As shown in above fig 4.7, Naïve bayes algorithm was unable to classify reviews according to their feature weightage. Consider a review "Battery is good excellent display great performance." In this review, display feature having more weightage or more impact than the battery and performance. But, naïve bayes algorithm is unable to classify particular review into display feature.

To overcome above drawback, we made some changes to naïve bayes algorithm. The modified naïve bayes shows better accuracy and performance.

To prove that modified naïve bayes shoes better results, same set of reviews are considered for modified naïve bayes algorithm.

Fig. 4.8 Result of reviews using Modified Naïve Bayes Algorithm

As shown in above fig. 4.8, Modified Naïve Bayes algorithm is successful in classifying the reviews according to feature weightage. Consider the same review "Battery is good excellent display great performance." In this review, display has more impact than the other features such as battery and performance. Modified naïve bayes successfully classifies this particular review into display feature. From the results, we can conclude that modified naïve bayes approach performs very well than simple naïve bayes approach.

Proposed system shows "Topical Relation" using FBC form. FBC form takes feature preferences from the user; system provides review and name of product which satisfies product feature preferences. For the experimental purpose, same dataset has been used.

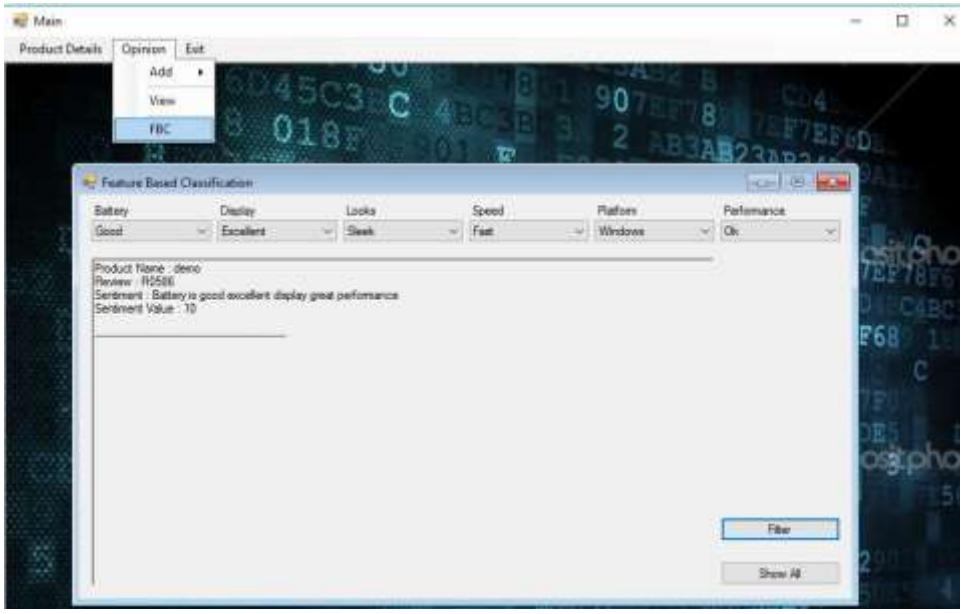

Fig. 4.9 Results obtained by FBC

As shown in above fig. 4.9 we select Good value for Battery feature, excellent for display, sleek for looks, fast for speed and ok for performance. In the output we get Nokia 660 product which contains all those features which are listed by users. This shows our topical relation between opinion target and opinion words.

4.3 Comparison of proposed system on different dataset

To show that, our system is compatible with other domain reviews, we took different dataset of "hotel reviews". These reviews are gathered from tripadvisor.com and makemytrip.com. After adding reviews, we added relevant features to extract the reviews; these features are "room" and "food". Results obtained by the system after adding new dataset are shown as below.

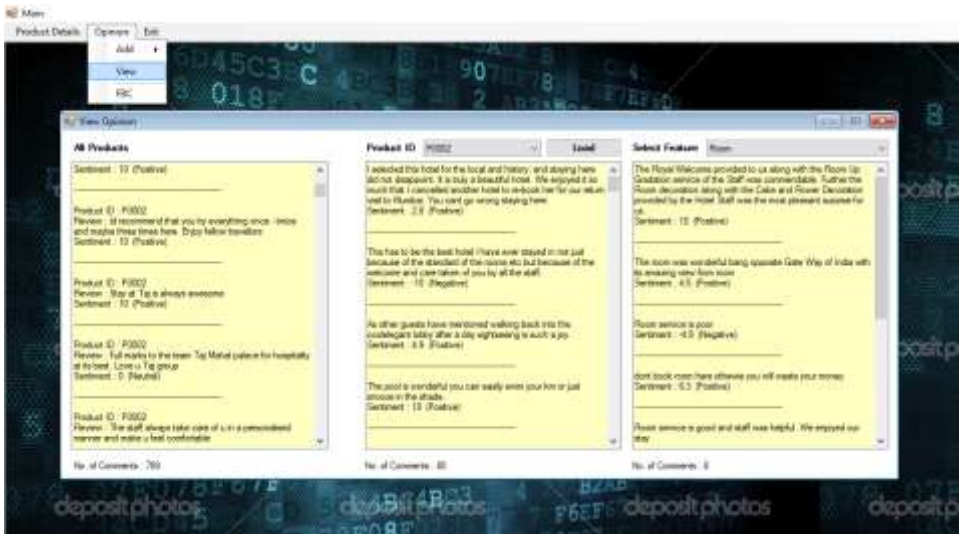

Fig. 4.10 Result of hotel reviews using Naïve Bayes algorithm.

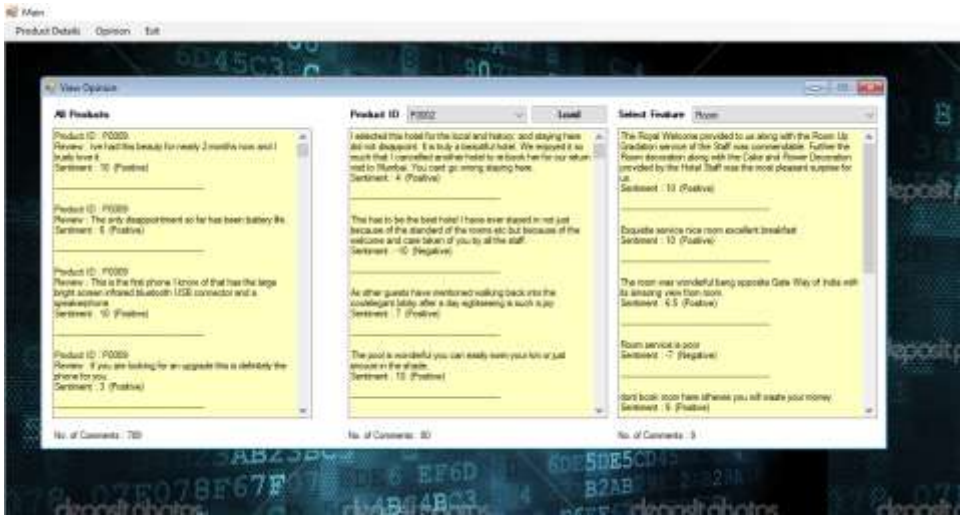

Fig. 4.11 Result of hotel reviews using Modified Naïve Bayes algorithm.

As shown in above fig. 4.11, this system is capable to classify the hotel reviews using modified naïve bayes algorithm. Modified naïve bayes algorithm gives better sentiment value than naïve bayes algorithm and classifies the hotel reviews in "room" feature.

From above experiment it is clear that, proposed system can run on various domain dataset.

4.4 Comparison of the algorithms

Here we have shown the comparative analysis between our two algorithms. To prove our proposed method is better than the simple naïve bayes, we took review data with different no. of records such as 100, 300, 500 and 1000 records. We consider five different performance measures to evaluate the performance of both algorithms.

Table 2. Comparison of algorithms based on Performance Parameters

Performance Measures	No. of Reviews							
	100		300		500		1000	
	NB	MNB	NB	MNB	NB	MNB	NB	MNB
Accuracy	0.6	0.71	0.62	0.62	0.54	0.59	0.52	0.55
Precision	0.79	0.92	0.71	0.87	0.63	0.82	0.58	0.72
Recall	0.81	0.88	0.63	0.69	0.67	0.72	0.62	0.65
Negative Predictive Value	0.43	0.6	0.4	0.63	0.3	0.58	0.26	0.3
Specificity	0.4	0.7	0.32	0.36	0.28	0.4	0.18	0.2
F-measure	0.8	0.9	0.67	0.77	0.64	0.76	0.59	0.68

From the above comparison analysis it is clear that, our proposed method gives better results than the simple naïve bayes irrespective of the no. of reviews.

V. Conclusion

The Opinion mining system has evolved significantly and will be evolving to provide the user with much efficient results. The system had aimed to classify the user reviews into their correct classes according to the features of the product. The simple naïve bayes algorithm is not able to classify review into particular class, if that review contains more than one feature mentioned. To solve this issue, we made some changes to naïve bayes algorithm. From experiment it is shown that modified naïve bayes correctly classify the reviews based on their highest probability of product feature. FBC form shows topical relation between opinion target and opinion words. Proposed system can be used to any other domains just by adding relevant features and review dataset of the particular domain.

Acknowledgements

I would like to take the opportunity to express my sincere thanks to my Guide Mrs. Harshali Patil, Assistant Professor, CMPN Department, TCET for her invaluable support and guidance throughout my P.G. research work. Without her kind guidance & support this was not possible. I am grateful to him for timely feedback which helped me track and schedule the process effectively. Her time, ideas and encouragement that he gave helped me to complete my project efficiently.

References

[1] A. M. Alkabani, A. M. Ghamry, F.K. Hussain and O.K. Hussain, Sentiment Analysis and Classification for Software as a Service Reviews, *IEEE 30thInternational Conference on Advanced Information Networking and Applications(AINA)*, Crans-Montana, 2016, pp.53-58, doi: 10.1109/AINA.2016.148.

[2] V. Dhanalakshmi, D. Bino and A. M. Saravanan, "Opinion mining from student feedback data using supervised learning algorithms," *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, 2016,pp.1-5. doi: 10.1109/ICBDSC.2016.7460390.

[3] K. Liu, L. Xu and J. Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 3, pp.636-650,March 1 2015. doi: 10.1109/TKDE.2014.2339850.

[4] H. P. Patil and M. Atique, Sentiment Analysis for Social Media: A Survey, *2nd International Conference on Information Science and Security (ICISS)*, Seoul, 2015, pp. 1-4. doi: 10.1109/ICISSEC.2015.7371033.

[5] Z. Z. Alp and S.G. Öğüdücü, "Extracting Topical Information of Tweets Using Hashtags," *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, 2015, pp.644-648. doi: 10.1109/ICMLA.2015.73.

[6] www.kavita-ganeshan.com

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Prajkta Akre. "Extracting Topical Relation between Opinion Target and Opinion Word using Modified Naïve Bayes Approach ." IOSR Journal of Computer Engineering (IOSR-JCE), vol. 19, no. 4, 2017, pp. 74–83.