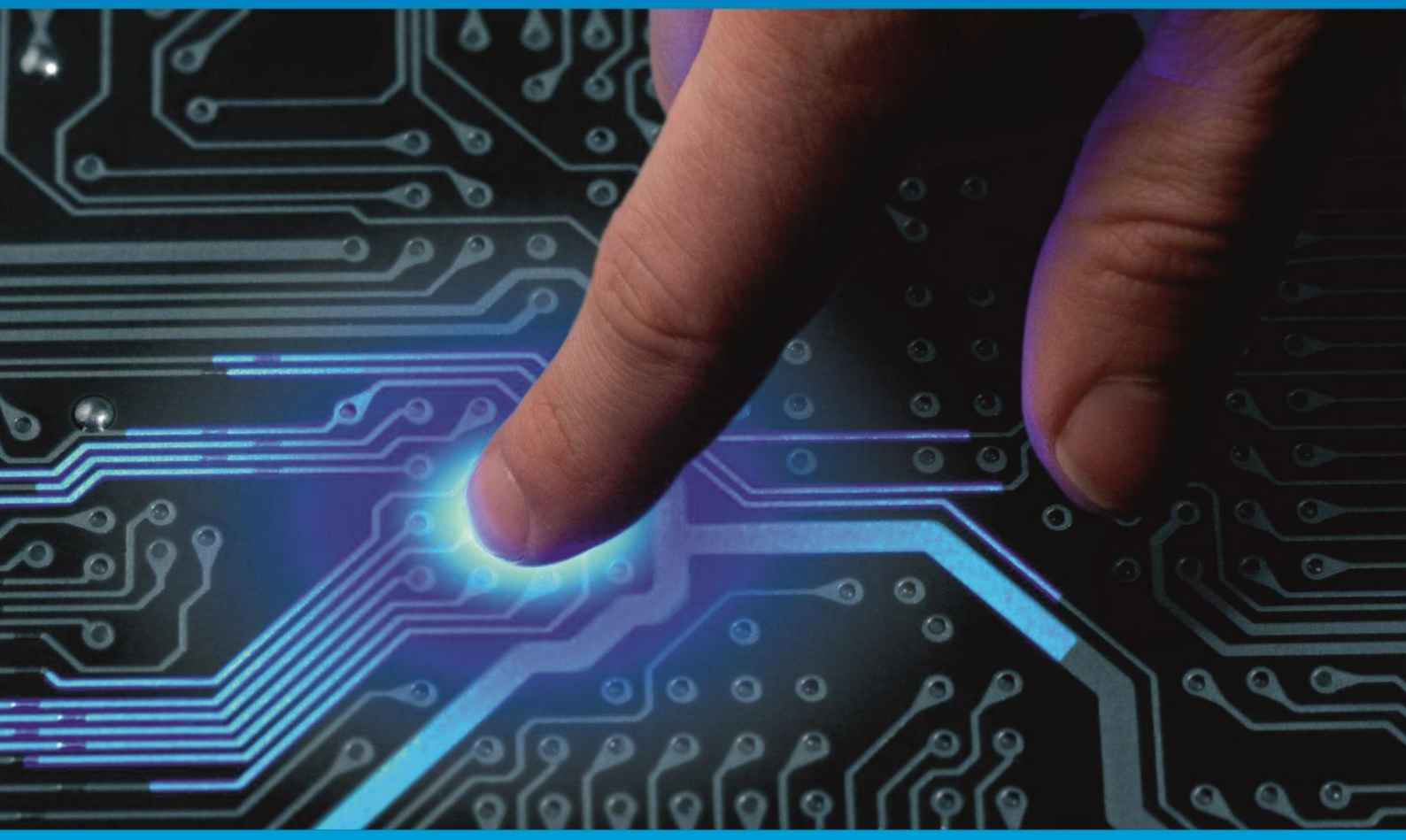
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 11, November 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 11, November 2020 ||

# Result analysis of Hybrid Feedback Analysis system for Educational Institutions

Nikita Dipak Bachhav, Prashant Nitnaware

PG Student, Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, India

Professor, Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, India

**ABSTRACT:** This paper gives a result analysis of an academic feedback system that allows students to enter their opinions anonymously. This automated feedback system collects opinions as a set of normalized and un-normalized responses collected by the users. This system gives sentiment analysis which computes positive and negative opinions from available feedbacks followed by a Naive Bayes and SVM classification algorithm to refine the analytical results. The conclusion is based on combined values from all these analysis. This proposed system significantly help to improve the quality of service of educational institutions by providing accurate results.

**KEYWORDS**: Supervised learning, Unsupervised learning, Sentiment Analysis, Machine Learning,natural language processing

## I. INTRODUCTION

Feedback is an important aspect to improve quality of product and services. Increasing high use of internet many ways has formed to acquire these feedbacks from the consumers in any form such as reviews, polls, comments, and rating. Various e-commerce platforms collect opinion of their active consumers regarding their product and services. Almost every organization is using forums, social media and online platforms to allow its consumers to provide their experience/advices regarding the commodities. These opinions can provide information about user's choices, likes, purchases and many other things. Through feedbacks, companies not only improves their quality of services but also identifies the ways of new service recommendations for their consumers.

Feedback can be categorized as normalized and un-normalized responses. Normalized responses are the responses given in the form of direct answer, for example, objective questionnaire or yes/no response. Un-normalized responses includes subjective expressions and complete description of experience [1]. Feedback contain useful information but it's difficult to extract the actual sense of each feedback accurately through automated system. Many techniques have been developed to extract the emotions from the text or images to know what exactly user wants to say.

We are using many machine learning algorithms for discovery of accurate prediction of features of data without external interference. The training data set predicts the feature of testing data set and concludes the result. Based on the availability of training data, the learning algorithms have been classified as supervised and unsupervised learning methods.Supervised system is trained using training data set wherefeatures are already known and based on these features training data identifies the features of testing data and categories them accordingly. On the other handunsupervised learning methods are useful when features are unknown.

Proposed system includes Naive Bayes classification technique which is a supervised learning mechanism which uses very small set of training data for estimation of features required for classification. Similarly, SVM classifier is again a type of supervised learning model which uses large set of training data. Here Reviews are classified at sentence level. Each sentence is analyzed separately and classified as negative, positive or neutral. Thus, overall document has a number of sentences where each sentence has its own polarity which mainly calculated using sentiment analysis, Naïve Bayes and SVM Algorithm.

## II. RELATED WORK

In [2] authors proposed a framework which analyzes student feedback. In thisconceptual framework, natural language processing are used for pre-processing and text mining techniques are used to extract the different dimensions from responses where correlation is performed to link the various topics within qualitative feedback. This will help for deep analysis of the collected data. Inequal manner [5] proposed a system using combination of lexicon based approach and machine learning techniques for sentiment analysis of student's feedback which is trained using TF-IDF and lexicon based featured to analyses sentiments by students expressions. In [8] authors has developed a system to analyses the students feedback which was collected from Social network and applied a machine learning algorithm likeMaximum entropy, support vector machine, naive Bayes and complement naïve Bayes to classify sentiments which define the optimal tool for feedback analysis. In [9] have proposed a framework for review analysis of onlineBox

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 11, November 2020 ||

Office movie reviews which used both clustering and classification. They have mainly focuses on unsupervised clustering algorithms like DB-SCAN and K-Means but system is unable to handle the double negation. In [12] authors have proposed a combined rating system to address the issue of numeric rating and opinion analysis by combining the numeric polarity of starred rating and reviews. In [10] authors has developed a framework to analyze the product review of ecommerce site. It has been filtered noisy data and has been pre-processed to evaluate sentiment of the reviews using supervised algorithms like Support Vector Machine (SVM),Naïve Bayes and Decision Tree. In [13] authors have proposed a framework on mining review from e-commerce websites uses Naïve Bayes text classifier, logistic regression. In this paper, SentiWordNet used to classify the review which concludes that based on their results Naïve Bayes is efficient among three algorithms. In [16], authors have proposed a feedback system for ecommerce organization which will help buyer to select best product on basis of analysis performed. It will collect data from e-commerce sites and done Pre-processing on collected reviews. Comments will be classified in three classes negative, positive and neutral. Features will be identified to calculate the trust score. In [18], [21] authors has developed a system to classify product review in semantic meaning. This system works with variant approach including spelling correction and implemented a hybrid algorithm using Naïve Bayes and Decision tree. Data has been collected using web crawlers and pre-processing is done to reduce ambiguity, redundancy and incomplete data (reviews). SentiWodNet is used to assign a polarity to each word. Naïve bayes is used to classify the data and decision tree calculates the final polarity after classification. Main aim of these systems is to remove anaphora's occurring in a review. To remove this occurring reference resolution is used. SentiWordNet is used to assign the scoresand Stanford parser is used to extract the opinion. In [19], [20] authors has discussed the work on sentiment analysis based on supervised machine learning approach but rule based classifiers are not used much. According to the authors, unsupervised techniques are useful when it comes to complicated training data set compare to rule based approach. Reviews are the most used form of data sources available widely. Opinion mining plays an extreme important role while making a decision towards a particular product.

## III. PROPOSED ALGORITHM

This work proposes simple but efficient algorithm for feedback analysis for an educational institute. Some of the limitations were found in the existing feedback analysis system. They are listed below:

- Most of the works are basedon unstructured data for feedback analysis. This data gives the polarity of positive and negative words.
- In qualitative feedback analysis, some systems only performedcorrelation to link the various topics.
- Most of the work are limited to compute only overall feedback of students.
- Very few paper are intent to get features to derive depth analysis of provided feedbacks.

In order to overcome these drawbacks, a hybrid feedback analysis system is proposed [1]. The proposed system is consist of mainly 4 phases

- Data collection phase
- Sentiment Analysis
- Classification
- Report Generation and Visual representation

Admin has full privileges to add or remove the names of any faculty from the list of feedback analysis system. Admin can also add or remove the subjects and programs offered.Admin can download the tabularreports in pdf, word format. The different phases of the system are as follows.

### A. *Data collection phase:*

In this phase, studentsneed to select the program they areenrolled in. After selecting a program, courses list will be displayed. Students will add feedback for their respective courses and teachers. The feedback is in normalised as well as un-normalised form. Students need to give their feedbacks marking their responses from the given options.The questionnaire along with a text box is given, so that, they can express their opinions in their own way about other factors which are not included in the objective questionnaire. Feedback will be anonymously submitted to the system with session id. On Similar basis, there is another module where facultiescan provide feedback regarding the courses

.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 11, November 2020 ||

B. *Sentiment analysis:*

In order to process un-normalised data sentiment analysis need to be performed. Input isprogramme and subject, once the admin will select a subject and program it will fetchdata from database for that particular subject. This phase is consist of two parts. One ispre-processing and another is sentiment analysis which will calculate the polarity.To extract useful information fromthe unstructured text, several pre-processing steps are applied to remove spelling errors,grammatical mistakes, URLs, etc. from the text[1].

Step 1: Sentence Segmentation

Student feedback data represents an unstructured text. Here First step is to identify each sentence. Sentence is segmented by identifying the boundary of sentence which ends with full stop.

Step 2: Tokenization

Tokenization is the process of breaking up a sequence of strings into pieces such as words,keywords, phrases, symbols and other elements words by identifying the comma, spaces and special symbols between the words called tokens. Tokens can be individualwords, phrases but not a whole sentences. In the process of tokenization, some characterslike punctuation marks are discarded [7].

Step 3: Stop word removal

Stop words are used in every language. We need to only consider the words in the document which have importance to effectively use word feature score.Stopwords are eliminated to focus on important words. For example, search engine query is “Who is thePresident of India.”, in this type of query search engine searches for the words like“who”,” President”, ”India”.Then it retrieves more pages containing thewords “India”, “President”. So, by deleting or removing these stop word we can get proper data for analysis.

Step 4: Stemming

In this step, the words are reduced to its base form or root word which is known as stem. Stemming is a mandatory pre-processing step in number of natural language processing applications like word sense disambiguation. One of the widely used tools for this processing is stemmer which keep a suffix list to remove suffix from its word. Porter stemmer which is one of the popular stemming algorithm is a process for removing the commoner morphological and in flexional endings from words in English. It removes prefixes and suffixes to get the affix free word by porter rules

Step 5: Polarity measurement

In sentiment analysis, it will calculate the overall number of count for comments. Afterpre-processing unstructured data tokens will be collected. Sentiwordnet is lexicalresource for opinion mining. Sentiwordnet assigns scores as positive, negative and neutral. Score for each token will be calculated with the help of Sentiwordnet.

C. *Classification:*

This classifier uses Naïve Bayes theorem to predict the probability, that a given set of features is a part of particular label. It uses bag of words model for feature extraction. This model assumes that all the features are independent[18].

$$P(\text{label/features}) = P(\text{label}) * P(\text{features/label}) / P(\text{features})$$

where P(label)= prior probability of label

P(features/label) =prior probability that feature set is classified as label

P(features)= prior probability that feature set will occur.

A Naïve Bayes classifier implements Bayes theorem with simplified assumptions toclassify inputs based upon the presence or absence of features. Here, the features aredistinct words and the classifier characterizes each training data comment. Using thischaracterization, the classifier computes the most probable category of each newcomment based on the presence or absence of words form the training data. The NaïveBayes classifier implementation used here is classify, which is designed as a multipurposeclassifier for the classification of text sources. It is designed to be quick andsimple to use – and offers no further control other than input texts. No manual pre-processingof the text was done before loading into the classifier system. As before, setcomments were used to train the classifier for positive and negative sentiment. Someindividual test data comments will be loaded into the classifier, and the category with thehighest probability will be recorded [18].As the comments will be positive, negative and neutral. There are only two predefinedclasses that is positive and negative. As the new comments will come classifier willcalculate its probability whether that comments is positive or negative.

Here another classifier, Support vector machine anon-probabilistic classifierengine is used to extracts the features. Support vector machine is a supervised learning model which is used for classification. Its main aim to determine best

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 11, November 2020 ||

linear separators for classification. In student's feedback system, normalised data collected from students isprocessed with the help of support vector machine engine. Support vector machine willextract the feature from normalised data and calculate the sentiment value for thatfeature.Factors like Gender, Session id, program, opinion, and its Sentiment value plays an important role in analysis.We also introduced some factors like male, female, in favour of faculty, against faculty, in favour of faculty change, not in favour of faculty change, overall probability which are the important keys of our project. List of questions will be given to students from that questions feature will beextracted [16]. For example,

Did the course content meet your expectation?
Feature- Course outcome
Course increases your knowledge?
Feature- Learning
Where the lecture easy to understand?
Feature- Explanation methods
Did the teacher explain the material well?
Feature- Teaching skills
Course Completion in Time?
Feature- Time
Standard of Lab facilities?
Feature- facilities

In this way features will be extracted from the questionnaire. Now the count of positive andnegative will be displayed for each feature after support vector classification. It willclassify data linearly. In this phase it will give the output for features and their respectivepositive negative scores. Also analysis of feedback will be displayed in this, which willinclude session id, course, and sentiment value per comment wise. Also the detailedanalysis of the feedback can be reviewed in the analysis section [1].

D. *Report*

This the last phase of hybrid feedback analysis system. In this phase, the detailed tabular reports will be displayed in the form of word/pdf which can be downloaded for further use. This report will include the polarity of the opinion i.e. positiveand negative output of sentiment analysis, refined results of Naive Bayes classification module and feedback analysis using support vector machine. Admin candownload these reports either in word or pdf and forward it to higher authorities. Along with the tabular report analysis, results could be displayed in graphical format for better visual representations and comprehension which improves the overall experience and Understanding.

## IV. PSEUDO CODE

Step 1: Take responses from Student in both Normalized and Un-Normalized form.
Step 2: Preprocess the reviews to measure the polarity
Step 3: Check the below condition for each reviews is available
if (Covered by Sentiment Analysis)
    if (Factors Identified)
        Report to Admin
    else
        Apply SVM classification
else
    Apply Naive Bayes classification
end
Step 4: Calculate the polarity using Sentiment Analysis, Naive Bayes and SVM classification.
Step 5: Generate the report in tabular format according to the response given by system.
Step 6: Calculate the result analysis in graphical format.
Step 7: End.

## V. SIMULATION RESULTS

The experiments were taken on a standalone machine with high storage capability. Machine is equipped with Intel core i5 processor having a clock speed of 2.4 GHZ and 8GB of RAM. We have to simulate the system which done the process of polarity measurement, using Sentiment Analysis, Naive Bayes and SVMclassification, report and graph summary. By looking at the results provided by proposed algorithm and the results obtained by studying different feedback summary modules, which compared against various factors and following points are worth noted.

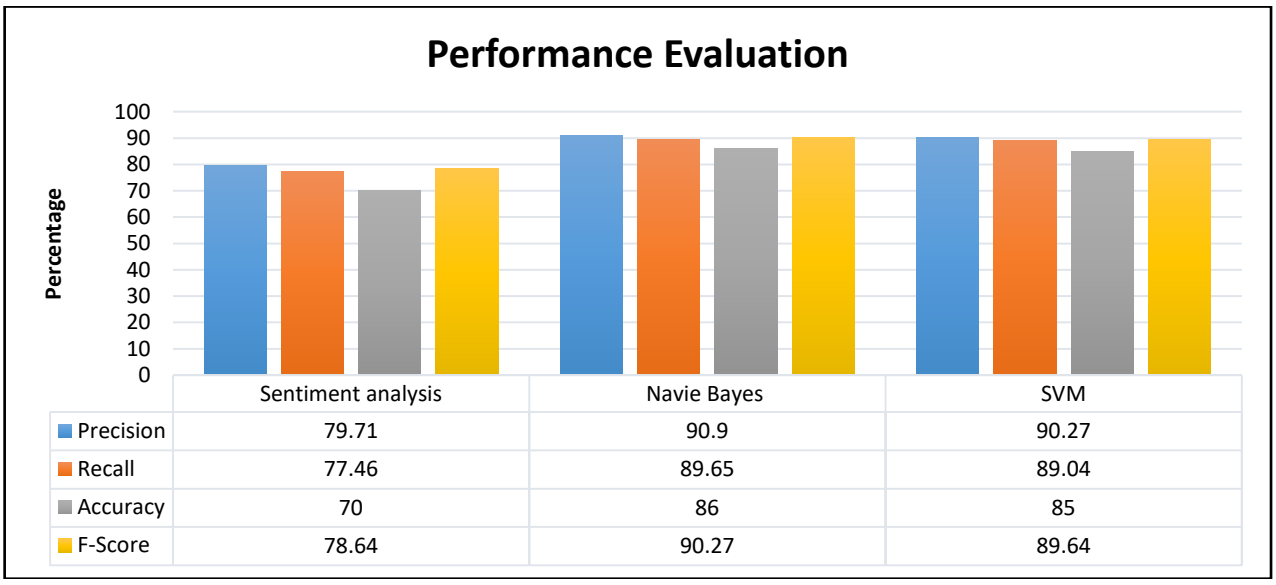
For accurately evaluating our system, there is a need of metrics which measure the actual performance of our system. In any polarity measurement solution, it is very much important to retrieve near precise answers which will be useful to

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 11, November 2020 ||

the user to understand.To measure the quality of polarity measurement solution, precision and recall are used as most common measures. Recall provides percentage of polarity calculated whereas precision provides fraction of all polarity calculated correctly. Recall provides correctness of system.Precision specify exactness of system. In fact, high precision regarding a system indicates that system extracted more relevant answers than irrelevant whereas high recall indicates system extracted most of the relevant answers present in the system.

Precision and recall can be easily defined using a contingency table where TP indicates number of true positives; FP indicates number of false positives, TN indicates number of true negatives and FN indicates number of false negatives.




| | Sentiment analysis | Navie Bayes | SVM |
|---|---|---|---|
| Precision | 79.71 | 90.9 | 90.27 |
| Recall | 77.46 | 89.65 | 89.04 |
| Accuracy | 70 | 86 | 85 |
| F-Score | 78.64 | 90.27 | 89.64 |

Fig.1.Opinion Mining of Student reviews using Precision, Recall, Accuracy and F-Score

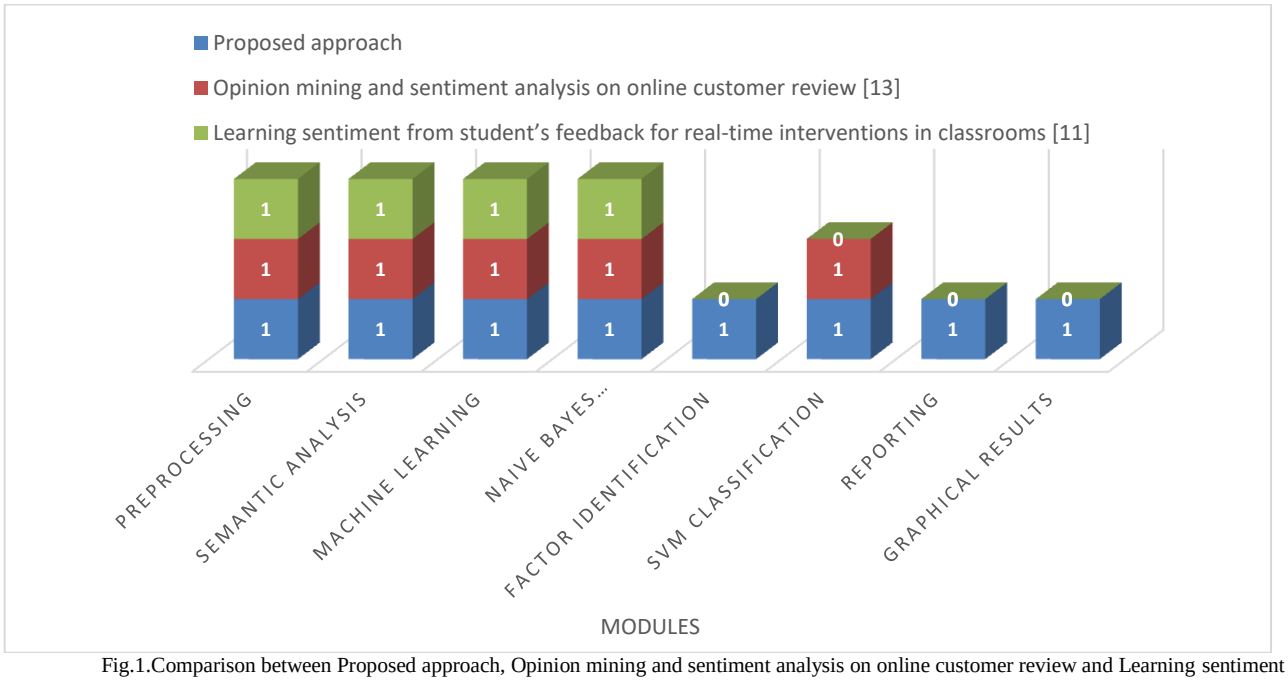



Fig.1.Comparison between Proposed approach, Opinion mining and sentiment analysis on online customer review and Learning sentiment from student's feedback for real-time interventions in classrooms

Above figure shows the Comparison between proposed approach, sentiment analysis, Naïve Bayes and SVM classification machine learning algorithm of Student reviews using Precision, Recall, Accuracy and F-Score. This shows clearly that the proposed system is much efficient. Also both machine learning algorithm are much efficient on small set of dataset.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 11, November 2020 ||

Comparisonbetween Our Proposed Approach, Opinion mining and sentiment analysis on online customer review [13] and Learning sentiment from student's feedback for real-time interventions in classrooms [11]. Number of Modules increases the accuracy of the system. Multiple factors in the opinion mining, where the reviewer has been provided. First we consider the points which covers the features and granularity of reviews to minimize the uncertainty which is covered in our proposed system.. These factors are provided by reviews to demonstrate their desires. Where the other models hardly include all these factor which will leads for reduced overall accuracy.

## VI. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed systems data mining methodology for feedback analysis helps to understand the performance of the faculties of an educational institution based on qualitative as well as quantitative feedback. Processing the feedbacks submitted by the students using sentiment analysis, Naive Bayes and support vector classifier gives accurate and efficient results which helps to maintain or improve higher educational standards. This system can be used to collect feedbacks for other organizations with related set of questionnaire. We have used very small set of 1000 reviews, as number of reviews increases the complexity will increase. We can increase the number of reviews and analyze the performance.

## REFERENCES

1. Nikita DipakBachhav, Dr.PrashantNitnaware, "Hybrid Feedback Analysis system for Educational Institutions", International Journal of Innovative Research in Computerand Communication Engineering, Vol. 7, Issue 6, June 2019.
2. Sandy Gan, SwapnaGottipati, VenkyShankararaman, "A conceptual framework for analysing student's feedback," 2017.
3. M. Ferna´ndez-Gavilanes, T. A´ lvarez -Lo´pez, J. Juncal-Mart´ınez, E. Costa-Montenegro, and F. J. Gonza´lez-Castan˜o, "Unsupervised method for sentiment analysis in online texts," Expert Systems with Applications, vol. 58, pp. 57–75, 2016.
4. 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.
5. ZarmeenNasim, Quratulain Rajput and SajjadHaider, "Sentiment Analysis of Student Feedback Using Machine Learning and Lexicon Based Approaches," 2017.
6. L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexiconbased and learning-based methods for twitter sentiment anal- ysis," Technical Report, 2011.
7. O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to the sentiment analysis problem at the sentence level," Knowledge-Based Systems, vol. 108, pp. 110–124, 2016.
8. Mohammad AmanUllah, "Sentiment Analysis of Student Feedback: A case study towards optimal tool," 2016.
9. Nagamma P., Pruthvi H.R, NishaK,"An improved sentiment analysis of online movie reviews based on clustering for Box office,",2015.
10. ZeeniaSingla, Sukhchandan Randhawa, Sushma Jain, "Sentiment Analysis of 23 Customer Product Reviews Using Machine Learning,",2017.
11. N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Learning sentiment from student's feedback for real-time interventions in classrooms," in Adaptive and Intelligent Systems. Springer, 2014, pp. 40–49.
12. Mir Riyanul Islam, "Numeric rating app for google play store by sentiment analysis on user reviews," 2014.
13. Santhosh Kumar K L,JayantiDesai,JharnaMajumdar., "Opinion mining and sentiment analysis on online customer review,", 2016.
14. Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teacher"s evaluation," Applied Computational Intelligence and Soft Computing, vol. 2016, p. 1, 2016.
15. Han, H.J., Mankad, S., Gavirneni, N., &Verma,R.(2016).Whatguests really think of your hotel: Text analytics of online customer reviews. Cornell Hospitality Report, 16(2), 3-17.
16. SonaliBagul, RakhiWajgi, "Design Feedback Analysis System for ecommerce Organization." 2016.
17. Balaji. Jagtap and V. Dhotre, "SVM and HMM Based Hybrid Approach of Sentiment Analysis for Teacher Feedback Assessment" International Journal of Emerging Trends of Technology in Computer Science (IJETCS). vol. 3, no. 3, pp. 229–232.
18. Gurneet Kaur, AbhinashSingla, "Sentiment Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm", 2016.
19. Neetika Bansal, Ashima Singh, "A Review on Opinionated Sentiment Analysis based upon Machine Learning Approach", 2015.
20. Shivprasad T K, Jyothi Shetty, "Sentiment Analysis of Product Reviews: A Review", 2017.
21. Hari Krishna, Ali Akbar, "A Feature Based Approach for Sentiment Analysis using SVM and Co-Reference Resolution", 2017.

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details