SURVEY ON SENTIMENTAL ANALYSIS USING MACHINE LEARNING ALGORITHMS

[1]*D. Saraswathi*
Asst Professor of
PG-Computer Science
SaradhaGangadharanCollege
Puducherry

[2]*F. Jeno Sandana Brina Rouvier*
HOD, Asst Professor of
Computer Applications
SaradhaGangadharanCollege
Puducherry

[3]*B. Venba*
Asst Professor of
PG-Computer Science
SaradhaGangadharancollege
Puducherry

ABSTRACT

Feedback mining is a tool for analysing the learning environment. In educational institutions, the most standard procedure is used to collect feedback from students in examining their feelings toward the subject instructor in terms of improving the teacher's performance. Generally, this method of evaluation through such feedbacks begins at the end of the semester with the use of survey forms. Students provide input on a variety of topics based on their actions in the classroom, lab, and elsewhere. The lexicon-based approach and the machine learning algorithms are the two types of techniques used in sentiment analysis. A variety of classification techniques, namely decision trees, Naive Bayes, SVM, and neural networks, are used to categorize the feedback.

Keywords: Sentimental Analysis, Lexicon-based, Feature Selection, SVM, Naïve Bayes, Hybrid Technique.

I. INTRODUCTION

Sentiment analysis is a method of detecting positive or negative sentiment in text. It's commonly used to detect sentiment in social data, analyze brand reputation, and develop a deeper understanding of their customers.

The act of converting unstructured text into a structured format in order to find new insights and patterns is known as text mining, also known as information mining. Advanced analytical approaches such as Nave Bayes, Support Vector Machines (SVM), and other deep learning algorithms can also be used to discover hidden relationships within unstructured data.

Sentiment analysis is performed on three levels: sentence level, document level, and features or aspect level. In general, sentimental analysis can be divided into three methods. Lexicon-based, machine-learning based and hybrid approaches. Unsupervised or supervised learning are both used in machine learning techniques. Several algorithms such as support vector machine, naive bayes, and random forest can be used to solve the classification problem. Sentiment polarity of textual content is used in lexicon-based methods.

Application of Sentiment analysis

- Online review classification
- Social media analysis.
- Students' feedback classification.
- Domain-specific applications.
- Email analysis
- Stock price prediction

II.   LITERATURE REVIEW

Brenan and Williams[1] stated that virtually all educational institutions collect various forms of feedback, such as teaching and learning methodologies, teaching materials, and evaluations, in order to improve the quality of education provided. The feedback results can be used to help improve the current learning process and ensure that learning objectives are fulfilled.

The concealed sentiments in the reviews could be classified as positive or negative.[2] It's a major challenge that combines data mining and natural language processing in order to obtain opinions or attitudes. [3] Data mining is increasingly used in educational settings, in addition to strengthening educational quality. [4]

Sentiment Analysis, also termed Opinion Mining, is a Natural Language Processing and Information Extraction activity that determines a user's sentiments or perspectives, which are expressed in the text as positive, negative, or neutral remarks and quotes. Naive Byes, Maximum Entropy, and SVM are examples of supervised or data-driven sentiment analysis techniques. A support vector machine (SVM) is used for sentiment classification, which measures sentiment classification accuracy [5].

To achieve the study objective, more input components were obtained for analysis, and the feedback collection format was modified. Machine learning algorithms (Nave Bayes and Support Vector Machine) were used for sentiment categorization, while the NRC Emotion Lexicon was used for emotion analysis. [6].

Input dataset comprises student feedback collected through the use of a Google form. After receiving test instances, the trained program uses machine learning methods to categorize the text as negative, neutral, or positive. Data preparation, student feedback, feature extraction, model training and test evaluation, and graphical representation are the six components of this work. Random forest, multinomial Nave Bayes Classifier, Support Vector Machine (SVM), and multinomial Nave Bayes Classifier also were used for the model training. A graphical presentation of the result outcome. [7]

2.1 Lexicon-based approach

A lexicon is a collection of words connected with polarized sentiments. The pre-processing of the text to be examined is the first step in the lexicon-based approach. Then it is determined whether the lexicon is contained in the dictionary of terms and, if it is, whether it is positive or negative, and the score is changed accordingly. As an outcome, the ultimate score will determine whether the text is good or negative. Using lexicon-based sentiment analysis, students' textual feedback comments are used to judge the efficacy of teaching performance.

To identify the lexical polarity of words, a sentiment database in English words was constructed. In order to get a better result, the opinion is stored in our emotion word database, which contains terminology related to the academic sphere. Using their sentiment word lexicon, which includes academic opinion words, the authors proposed a method for evaluating teaching by classifying students' comments into seven categories (strongly negative, moderately negative, weakly negative, strongly positive, moderately positive, weakly positive, and neutral). [8]

This article cites opinion mining to collect feedback from students, which can then be evaluated to use a lexicon-based methodology for sentiment analysis.[9]

To find out fake reviews then follow process. The hybrid approach combines machine learning and lexicon-based techniques. The training data set is tagged with the sentiment lexicon, which is then used to train the machine learning model. The testing data is then analyzed using model. [10]

[11] The author used the VADER emotional analyzer tool, and also decision tree and nave bayes machine learning techniques. Dots are utilised to split the student review into two, three, and more sentences using the VADER emotional analyser. VADER and ML accuracy are compared.

2.2   Machine Learning
2.2.1 Feature selection methods

In order to provide an effective blended learning environment, it is useful to extract users' views or comments on this learning process. This work shows the experimental results achieved after comparing various feature selection approaches, namely Chi-square, Information Gain, Mutual Information, and Symmetrical Uncertainty [12].

Feature selection is a machine learning tool for extracting subsets of relevant features from a big dataset. As a result, the output quality is not affected when this method is used. The dimensionality of the feature space

is minimized when feature selection methods are used to classify data in text format. Applying feature selection to identify relevant features, this high dimensionality feature space reduction is achieved. [13]

The authors created a new technique for sentiment analysis using their SENTA tool, an open-source tool for extracting text features, in which tweets may be classified into seven categories (happy, sad, angry, love, fun, hate, and neural). The technique achieves precision of 60.2 percent and 70.1 percent after discarding all neutral tweets.

Feature selection algorithms can be useful in a wide variety of applications, including image processing, biology, banking, text categorization, and many others. The separation of different features is the basic notion behind this approach.

2.2.2 Machine Learning Algorithms for Classification

The dataset was developed to test the polarity of terms such as positive, negative, and neutral that students used as feedback. Afterwards, a Recurrent Neural Network (RNN) model was constructed for text categorization, which is based on the Long Short Term Memory (LSTM) approach for predictive opinion mining.[14]

The main focus of this research is on data cleaning and similarity identification, with algorithms likes K-Means, Latent Dirichlet Allocation (LDA), Scalable Multi-stage Clustering (SMSC), and GLIC are used to improve the degree of similarity for accuracy and comparison performance.[15]

The author works with different kinds of datasets: linear and non-linear datasets. A linear dataset has equal properties, whereas a non-linear dataset has properties that are not equal. The initial phase is data preparation, followed by feature extraction and classification algorithms like as Naive Bayes, ANN, and SVM.[16]

The researchers explain how to acquire academic domain data, then the preprocessing stage, the skip grams model for creating domain word embedding, and subsequently the working mechanism of our two-layer LSTM neural network for aspect extraction and sentiment classification.[17]

To improve accuracy, the author proposes using LSTM with attention layer, which combines sentiment words with embedding. Multi-head is a new neural machine technology that also employs the encoding method. To enhance the accuracy, the article suggested an LSTM+ Attention Layer+ Multi-head method model.[18]

In general NLP, the undesired text is first removed from the collected feedback. To increase feature selection, the following methods were used: bag of words, N-gram, POS, TOS tagging. Many predictive approaches are used to forecast the accuracy of comparisons.[19]

Sentiment analysis using the Teaching Senti-sentiment Lexicon's scores. The focus of this research is to use the suggested method for sentiment classification to automatically gather feedback from open-ended questions, as well as to determine the best classification of responses to open-ended questions within educational attitude settings by classifying attitude of positive or negative.[20]

2.3 Hybrid Technique

An hybrid algorithm for extracting sentiment from documents, is proposed by the authors. It is named as ASSAY. They used Nave Bayes and Support Vector Machine methods (machine learning algorithms) to identify the feedbacks of each domain, as well as HARN's algorithm, which is a lexicon-based approach to extracting the sentiment of a given document.

Table 1: A review of sentimental analysis utilizing machine learning techniques throughout the last few years

S.No	Paper Title	Author	Methods	Description	Accuracy	Year
1	Improving Teaching and Learning Experience in Engineering Education using Sentiment Analysis Techniques.	Wandeep Kaur, Vimala Balakrishnan, Baljit Singh	Data Processing Tokenization, stemming, normalization Classification Naïve Bayes (NB) and Support Vector Machine (SVM)	To examine the sentimental and emotional content of student feedback	70%	2019
2	Sentimental Analysis of Student Feedback using Machine Learning Techniques	Daneena Deeksha Dsouza, Deepika, Divya P Nayak, Elveera Jenisha Machado, Adesh N. D.	Support Vector Machine, Multinomial Naïve Bayes Classifier, and Random Forest.	Using many techniques in sentiment analysis to analyze student feedback.	80%	2019
3	Sentiment Analysis of Students' Comment Using Lexicon Based Approach	Khin Zezawar Aung, Nyein Myo	Lexicon based approach	Evaluating the level of teaching performance from students' textual feedback comment.		2017
4	Opinion mining technique for developing student Feedback analysis system using lexicon-based Approach(Opinion Mining feedback)	Muslihah Wook, et al.	Lexicon-based approach	Analysis of students' freestyle comments to evaluate instructor performance.		2019
5	Student feedback sentiment analysis system for distance education using arm with k-means clustering	G.N. Harshini and N. Gobi	Fake review indicators, Assosciation rule mining with k-means clustering, Lexicon method	Fake reviews can be discovered and removed from the dataset, and both implicit and explicit attribute extraction from the review sentence, as well as its polarity, may be identified using a Lexicon-based method.		2020
6	Sentiment Analysis of User Feedback in E-Learning Environment	Mohd Asri Omar, Mokhairi Makhtar, Mohd Fauzi Ibrahim, Azwa Abdul Aziz	Lexicon-based approach, Naïve Bayes and Decision Tree.	Data gathered from a prior research assistant and techniques based on machine learning and lexicon	90%	2020
7	Sentiment Research on Student Feedback to Improve Experiences in Blended Learning Environments	R.K. Kavitha	Feature selection methods say Chi-square, Information Gain, Mutual Information and Symmetrical Uncertainty. Classification; Decision Tree, SVM, Naïve Bayes classifier, k-NN	Both feature selection and classification were used in this study.	84%	2019
8	Student Feedback Mining System Using Sentiment Analysis	R.Menaha R.Dhanaranjani, T.Rajalakshmi, R.Yogarubini	Keyword Extraction , Topic Extraction Clustering, Classification	The pre-processed document is used to extract features or topics. Each topic's feedback remarks are gathered and organized into a cluster.		2017

9	Sentiment Analysis of Students' Comment using Long-Short Term Model	Irfan Ali Kandhro, Shaukat Wasi, Kamlesh Kumar, Malook Rind , Muhammad Ameen	LSTM,RNN	To increase accuracy.	90%	2019
10	Clustering Based Sentiment Analysis Using Randomized Clustering Cuckoo Search Algorithm	Samar H. Ahmed , Khalid Tawfik Wassif Emad Nabil	K-Means, Latent Dirichlet Allocation (LDA), Scalable Multi-stage Clustering (SMSC), and Grouping Like-minded people using Interests Centers GLIC Algorithm	The proposed architecture is made up of three primary components: data cleansing, similarity detection, and randomised data. Cuckoo search clustering (RCCS). A formula combining the degree of similarity is offered as a way to increase accuracy.	91.16%	2020
11	Performance Analysis on Student Feedback using Machine Learning Algorithms	Sharnitha Katragadda, Varshitha Ravi, Prasanna Kumar, G. Jaya Lakshmi,	Naïve Bayes Support Vector Machine Artificial Neural Network	Numerous evaluation standards for diverse methodologies have been developed.	88%	2020
12	Aspect-Based Opinion Mining on Student's Feedback for Faculty Teaching Performance Evaluation	Irum sindhu , sher muhammad daudpota , kamal badar, Maheen bakhtyar, junaid baber et al	Two layer LSTM neural network for aspect extraction and aspect sentiment classification.	Automatically collecting aspects and their respective orientations from the provided text.	91%&93%	2019
13	Sentiment analysis of student feedback using multi-head attention Fusion model of word and context embedding for LSTM	K. Sangeetha · D. Prabha	LSTM with attention layer Multi head attention Fusion model	Over a popular Natural Language Processing approach, combining multi layers with LSTM improves the result.	93.02%	2020
14	Student feedback prediction using Machine learning	Nilesh Singh Sengar et al	SVM, MLP, Decision Tree, K-star, Bayes Net, Simple Logistics, Multi-class Classifier And Random Forest. Ten-fold cross-validation (CV)	A set of classifiers is used to evaluate the performance of the student's predictive model.		2019
15	Sentiment analysis in teaching evaluations using sentiment phrase Pattern matching (SPPM) based on association mining	Chakrit Pong-inwong Wararat Songpan	Tokenization, SPPM with association , Feature selection	By categorizing attitude settings as positive or negative, it is possible to identify the optimal classification of responses to open-ended questions within educational attitude contexts.	87.94	2018

III. CONCLUSION

Opinion mining is very vital for any education institution for improving teacher performance. It is highly necessary to ensure that standard methods are applied to conduct sentimental analysis. This paper broadly elaborated various techniques that are used in sentimental analysis. A new technique known as the "hybrid technique," which combines machine learning with lexicon-based strategies to improve classification performance.

IV.   REFERENCES

[1]   R. Williams and J. Brennan, "Collecting and using student feedback Date: A guide to good practice," 2004.

[2]   Quan C, Ren F (2014) Unsupervised product feature extraction for feature-oriented opinion determination. Inf Sci 272:16–28.

[3]   Sun S, Luo C, Chen J (2017) A review of natural language processing techniques for opinion Mining systems. Inf Fusion 36:10–25.

[4]   Natek S, Zwilling M (2014) Student data mining solution–knowledge management system related to higher education institutions. Expert Syst Appl 41(14):6400–6407.

[5].   Venkata Rajeev P, and Smrithi Rekha V, Recommending Products to Customers using opinion mining of online product reviews and feature, 2015 International Conference on Circuit, Power and Computing Technologies.

[6]   Wandeep Kaur, Vimala Balakrishnan, Baljit Singh, Improving Teaching and Learning Experience in Engineering Education using Sentiment Analysis Techniques, ICAME 2019, IOP Conf. Series: Materials Science and Engineering 834 (2020) 012026, doi:10.1088/1757-899X/834/1/012026.

[7]   Daneena Deeksha Dsouza, Deepika, Divya P Nayak, Elveera Jenisha Machado,Adesh N. D. Sentimental Analysis of Student Feedback using Machine Learning Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019, pages 986-991.

[8]   K. Z. Aung and N. N. Myo, "Sentiment analysis of students' comment using the lexicon-based approach," IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017, pp. 149-154, 2017.

[9]   Muslihah Wook, et al. Opinion mining technique for developing student feedback analysis system using lexicon-based approach (Opinion Mining Feedback), Education and Information Technologies (2020) 25:2549–2560 https://doi.org/10.1007/s10639-019-10073-7, pages 2550-2560.

[10]   G.N. Harshini and N. Gobi, Student Feedback Sentiment Analysis System For Distance Education Using Arm With K-Means Clustering, Ictact Journal On Soft Computing, April 2020, Volume: 10, Issue: 03 DOI: 10.21917/ijsc.2020.0294, page 2071-2075

[11]   Mohd Asri Omar, Mokhairi Makhtar, Mohd Fauzi Ibrahim, Azwa Abdul Aziz, Sentiment Analysis of User Feedback in e-Learning Environment, International Journal of Engineering Trends and Technology (IJETT) – Editor's Issues – 2020, ISSN: 2231-5381,doi : 10.14445/22315381/CATI2P224,Pages 153-157.

[12]   R.K. Kavitha, Sentiment Research on Student Feedback to Improve Experiences in Blended Learning Environments, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-11S, September 2019.

[13]   R.Menaha, R.Dhanaranjani, T.Rajalakshmi, R.Yogarubini Student Feedback Mining System Using Sentiment Analysis, International Journal of Computer Applications Technology and Research Volume 6–Issue 1, 51-55, 2017, ISSN:-2319–8656, pg-51-55.

[14]   Irfan Ali Kandhro, Shaukat Wasi, Kamlesh Kumar, Malook Rind , Muhammad Ameen , Sentiment Analysis of Students' Comment using Long-Short Term Model, Indian Journal of Science and Technology, Vol 12(8), DOI: 10.17485/ijst/2019/v12i8/141741, February 2019, ISSN (Print) : 0974-6846 , ISSN (Online) : 0974-5645.

[15]   Samar H. Ahmed, Khalid Tawfik Wassif, Emad Nabil, Clustering Based Sentiment Analysis Using Randomized Clustering Cuckoo Search Algorithm, IJCSNS International Journal of Computer Science and Network Security, VOL.20 No.7, July 2020, pg:159-166.

[16]   Sharnitha Katragadda, Varshitha Ravi, Prasanna Kumar, G. Jaya Lakshmi, Performance Analysis on Student Feedback using Machine Learning Algorithms, 2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS)

[17]   Irum sindhu , sher muhammad daudpota , kamal badar, Maheen bakhtyar, junaid baber, and mohammad nurunnabi, Aspect-Based Opinion Mining on Student's Feedback for Faculty Teaching Performance Evaluation, VOLUME 7, 2019. Digital Object Identifier 10.1109/ACCESS.2019.2928872

[18] K. Sangeetha · D. Prabha, Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM, Journal of Ambient Intelligence and Humanized Computing (2021) 12:4117–4126 https://doi.org/10.1007/s12652-020-01791-9, pages 4118-4126

[19] Nilesh Singh Sengar, Kartik Chourey, Shraddha Bajaj, S. Abinayaa. Student feedback prediction using machine learning, International Journal of Advance Research, Ideas and Innovations in Technology, ISSN: 2454-132X Impact factor: 4.295 (Volume 5, Issue 2), www.ijariit.com

[20] Chakrit Pong-inwong Wararat Songpan, Sentiment analysis in teaching evaluations using sentiment phrase Pattern matching (SPPM) based on association mining, International Journal of Machine Learning and Cybernetics (2019) 10:2177–2186 https://doi.org/10.1007/s13042-018-0800-2

[21] D.V,N.Devi, T.V.R. Kanth, K.Mounika, N.S Swathi. "Assay: Hybrid Approach for Sentiment Analysis". In: Satapathy S., Joshi A. (eds) Information and Communication Technology for Intelligent Systems. Smart Innovation, Systems and Technologies, vol 106. Springer, 2019.