Sentimental Analysis on Student Feedback using NLP & POS Tagging

Nandakumar R
Department of Computer Science
Amrita Vishwa Vidyapeetham
Amrita School of Arts and Sciences
Mysuru, Karnataka, India
nandareddy.r1995@gmail.com

Pallavi M S
Department of Computer Science
Amrita Vishwa Vidyapeetham
Amrita School of Arts and Sciences
Mysuru, Karnataka, India
ms pallavi@my.amrita.edu

Pramath.P.Harithas
Department of Computer Science
Amrita Vishwa Vidyapeetham
Amrita School of Arts and Sciences
Mysuru, Karnataka, India
pramathharithas09@gamil.com

Vinayak Hegde
Department of Computer Science
Amrita Vishwa Vidyapeetham
Amrita School of Arts and Sciences
Mysuru, Karnataka, India
vinayakhegde92@gamil.com

Abstract— **Sentiment analysis has been extensively used in a variety of fields in the previous decade, including business, social media, and education. The usage of sentiment analysis is rising, but it remains difficult, particularly in the education area, where the nature of the language utilized by pupils and a sizeable amount of information, dealing with their processing thoughts is a difficult task. Several studies of the literature illustrate the current state of sentiment analysis application which include sector from various perspectives and circumstances. In the teaching analysis system, the qualifier sentimental words for student feedback aren't thought of the result of teaching analysis through student feedback isn't displayed whether positive or negative opinion. To efficiently utilize sentimental analysis for the teaching system through student feedback, this paper proposes to analyze the scholar's text feedback automatically using NLP and post-tagging based approach to conclude the extent of teaching work. Here, the main concentration is applying sentimental analysis for students' feedback, collected through online mode. The use of ML and the NLP approach to examine the feelings of the students included in the text document is used for classification based on students' comments to predict the polarity (positive/ negative). A total of 2200 student feedback are collected and analyzed, which results in 60 percentage positive and 40 percentage negative through Post Tagging.**

Keywords—**Natural language processing (NLP), Part-of-Speech (POS) Tagging, Student Feedback, Sentimental Analysis, Feedback Data set.**

I. INTRODUCTION

The current educational system is a landscape that is constantly enriched by a large amount of information produced every day in various formats, much of which hides useful and valuable information. Among the greatest benefits that sentiment analysis and opinion mining approaches may bring is the ability to find and extract invisible "pearls" from the vast quantity of educational data. Students' feelings and opinions are useful, not just for studying students' attitudes toward a stream, or subject, or teacher, including for improving education plans and foundation growth. Even if sentiment analysis and opinion mining appear to be exactly similar, there is a distinction, they formerly relate to

discovering sentiment words and phrases that express feelings, in extracting and analysing people's view for a certain organization. At the word sentence, or document level, sentiment analysis can be carried out for teaching analysis through student feedback. Manual sentiment handling, however, is impracticable due to the high amount of papers. As a result, data processing should be done automatically. NLP is used to do sentiment analysis from text-based, phrase or document-level corpora. Collecting student's feedback to analyze teaching system in academics plays a vital role for the growth of students, teachers and institution. The scholars give the feedback to convey what's the distinction between the particular teaching method that is presently going down in schools and what variety of teaching methods students except. These evaluations demonstrate the school's overall success in their respective subjects. This will help them to improve their teaching consequently by the method of gathering and processing knowledge for the effectiveness in teaching. Stu-dents used to give feedback using paper feedback forms in the previous system; it was an inefficient and time-consuming operation. The forms are then gathered, and the HOD (Head Of Department) reviews the student responses and evaluates the performance of that particular department's instructor. To solve the limitations of that system, online feedback systems were developed, which take student comments through online and analyze it automatically to evaluate teachers performance. However, current online feedback systems primarily assess objective type questions and ignore descriptive questions Sentimental analysis is a technique for determining the tone of a text. The importance of text sentiment analysis has grown in recent years as individuals throughout the world face more difficult situations. In sentimental analysis, there are three general approaches, lexicon-based, machine-learning, and hy-brid approaches.

Machine learning approaches can involve either unsupervised or supervised learning. The classification problem can be solved using a variety of algorithms like SVM, naive bayes, and random forests.

II. RELATED WORK

Kastrati [1] However, there isn't a systematic assessment of the research and outcomes of using natural language processing, DL, and ML solutions for sentiment analysis in the education area in the body of literature. [2] presented a dependency parsing-based method for identifying irregular letter's semantic orientation. The proposed method made use of sentiment lexicons de-veloped with a semi-automated polarity expansion algorithm. Sentiment analysis techniques that use supervised machine learning require training classifiers with linguistic information taken from the text.[3] described a supervised learning method for predicting sentiment from student comments. To train the describe models, the n-gram features extracted from the feedback text were employed. SVM, Maximum Entropy, and Naive Bayes techniques were used to train the models.The semantic intention of an opinion word was predicted using the online lexical database WordNet [4].Another lexicon-based strategy offered by [5] is to use the dictionaries created to regulate the polarity of a word.The application of lexicon to assess the sentiment polarity of a student's response was discussed by Rajput and Haider [6].In order to regulate the polarity of an academic view, they adapted a general-purpose sentiment dictionary. Their findings revealed that using a domain-specific sentiment vocabulary produced better results than using a universal sentiment lexi-con. To analyse Twitter data for [7] presented a hybrid technique. The training dataset was labelled with sentiment polarities using an opinion lexicon. On the evaluation dataset,To predict sentiment polarity, the tagged dataset was utilised to train a binary classifier. Appel [8] At the phrase level, we employed a hybrid technique to analyse sentiment. They employed a sentiment lexicon enriched with to regulate the polarity of a sentence's sentiment, SentiWordNet and fuzzy sets were used. In 2016, I noticed the Dictionary-Based Sentiment Analysis method. They talked about how sentiment analysis was done, as well as the obstacles and issues that came up during the process. SenticNet, SentiFul, SentiWordNet, and WordNet are some of the dictionaries used in sentiment analysis. Over a research website, lexicon-based approaches are cost effective.

Even if a generic lexicon such as WordNet is employed, the accuracy of the classifier is harmed by issues such as negation, synonyms, sarcasm, and so on. This has fueled significant growth in online shopping, making opinion analysis a critical issue for corporate development [9].
Another Approach was developed by Bhagyashree Gore supported Lexicon based mostly Sentiment Analysis of Parent Feedback to gauge their Satisfaction Level [10] in 2018. They used lexicon based mostly approach and computing of polarity values. Throughout this approach they produce a lexicon of words with opinion score assigned to that. Niranjan, B. S., Hegde, V. [11]A chatbot is currently being employed in the field of education. A chatbot can be constructed faster, easier, and more successfully by utilising techniques such as Natural Language Processing (, Natural Language Understanding , Pattern Matching Techniques, and IR. Hegde, V et al .[12] By surveying students and determining which kind of learning they prefer, the aim of this essay is to learn about their thoughts on online learning during the epidemic. It is performed by utilising the mean score to compare variables and a statistical method to determine correlation strengths by displaying them in bar graphs in R studio. Devasia, T et al.[13] The system, which keeps track of all student entrance records, stream information, subject information, student marks information, attendance information, and other information, tries to increase the success grid of students using Naive Bayesian. It uses the passed out students' as input and predicts both of the future performance on a semester-by-semester basis.

Hegde et al.[14]The information is gathered from under-graduate students through lab exams and teacher question-naires. Finally, the traits that are supported and those are not supported for predicting student success can be finished by taking into account student behavioural and computational thinking parameters and utilising basic linear regression. Hegde, V., Kamath, N.[15]Academics, demographical considerations, psychological concerns, health issues, teacher opinion, and student conduct all play a role in feature selection. In this research, we present an approach for predicting student dropout using the R language and the Naive-Bayes Classification Algorithm. [16] The government's prolonged total lockdown in reaction to the COVID-19 epidemic has prevented people from interacting physically. Social media communication significantly increased as a result of this. Design and evaluation of four language-based models for Nepali Covid-19 tweet sentiment analysis. Companies will profit from an AI-based sentiment analysis system because it is anticipated that the number of persons utilising social media will soar in the coming days. It will help businesses a great deal in their climate change adaptation. [17] When compared to earlier methods, this study provides a more effective and efficient automated feature extraction technique. Conventional approaches like the surface approach use the manual feature extraction procedure, which is a crucial step in feature driven advances. These

strategies provide a solid foundation for assessing the certain of the characteristics and are the ideal framework for incorporating deep learning methods.
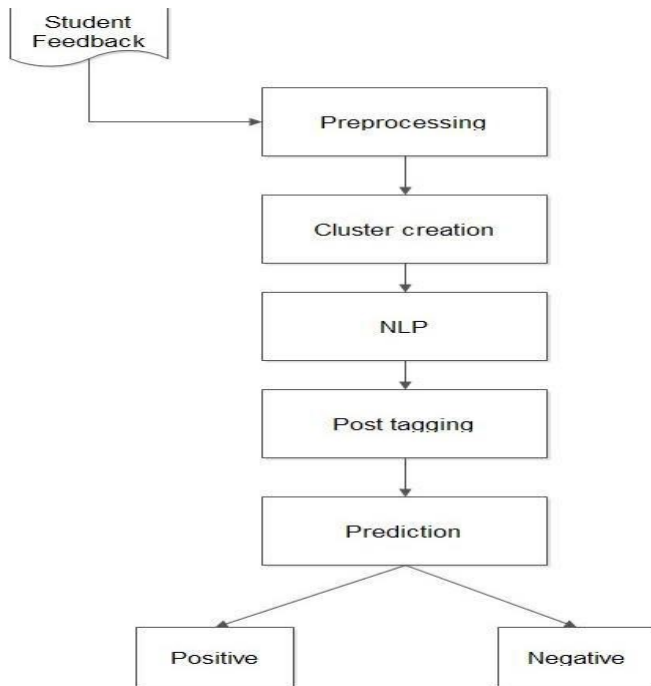
III. FLOW CHART

Fig 1: Proposed System Architecture

IV. METHODOLOGY

For sentiment classification, we use a machine learning-based approach in this system. We're doing this by creating a real-time dataset of student input. When the feedbacks are acquired, they are pre-processed to decrease noise. Each input is given a positive, negative, or neutral rating. From the feedback data set, meaningful and significant features are extracted after pre-processing. On the training dataset, a ML algorithm such as NLP is used. To test the model's accuracy, the model is applied to an unseen test dataset. The findings will be displayed at the end of a web application that has been constructed. For user convenience, the results are visualized and shown on a web application.

A. Data Collection
The dataset for this study is made up of 2848 comments taken from our institute's educational portal. Positive, negative, and neutral sentiment polarity labels were manually applied to the dataset. The data set has been collected from real-time students through an survey.

B. Data Preprocessing
There are three types of text pre-processing techniques:

1) Tokenization

Tokens make the student feedback in the text document. Which are groups of characters. The above text documents are divided into tokens, which are then used for data clarification.

2) Removal of Stop Words:
A stop-list or a collection of stop records may exist in an online search engine or other natural language processing system. "an," "a," "of," "the," "you," and "and" are the most regularly used stop words in English. This is a list of meaningless words. As a result, any words that appear too frequently and provide no information for the task are deleted.

3) PoS Tagging:
A POS tagger looks for parts of speech in a sentence or document and assigns them to each term. We utilised the Stanford POS tagger for POS tagging. This tagger breaks down text into sentences and assigns each word a POS tag. Consider the following scenario: "The staff is exceptional." Each word in the review is labelled using POS tagging (such as noun NN, adjective JJ, verb RB).

A 'movie' has the NN tag, indicating that it is a noun, whereas 'amazing' has the JJ tag, indicating that it is an adjective.

4) Feature Extraction
Student feedback are extracted from each lines throughout the feature extraction procedure. First, one must determine the polarity of a written document. comprehend the sentiment score, as well as its usage and relationship to all neighbouring terms. The following are some characteristics that influence the document's polarity.

5) Positive Emotion Words:
According to SentiWordNet, these are words with a positive sentiment score. Nice, Good, Fantastic, Pretty, Outstanding, and so forth.

6) Negative Sentiment Words:
According to SentiWordNet, these are words with a negative sentiment score. Bad, Awful, Disgusting, Pathetic, and so on.

7) Feature Reduction:
One of the most difficult challenges in sentimental analysis is dealing with text data with many dimensions, which can reduce the efficacy of the classifier. As a result, there is a need for a technique that will eliminate non-essential qualities while maintaining only the most cru-cial ones, as well as strategies that will aid to categorize phrases into positive and negative categories. Among the several feature reduction strategies, the most common are the Information Gain and Gain Ratio.

8) Final Prediction
In the final prediction the sentimental of the feedback will be given for each and every subjects. By making use of the python packages like matplotlib and plotly the feedback data can be represented in the form of line graphs.
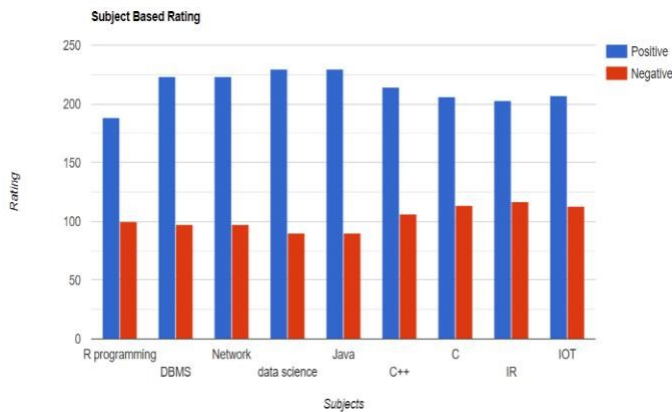
C. RESULTS AND DISCUSSIONS

Fig 2: Performance Graph

Performance graph shows the forecast for the feedback data set, the graph shows the sentimental analysis for each and every course. This was possible based on the students feedback for the particular course.
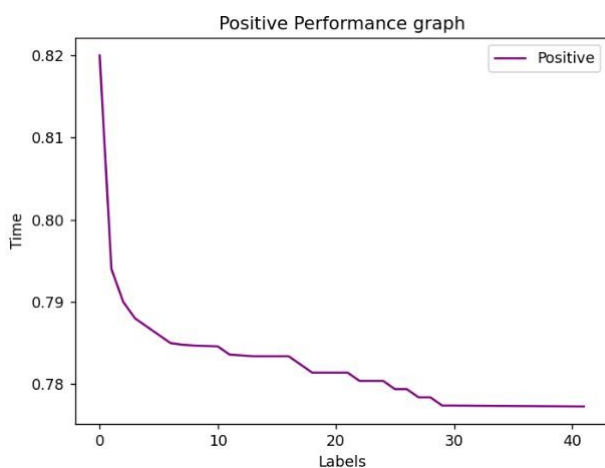

Fig 3: Positive Performance Graph

The above graphs shows how the model will be processed for the positive feedback. The time to build will decrease as the epochs is increased performance will be more to predict the positive feedback.
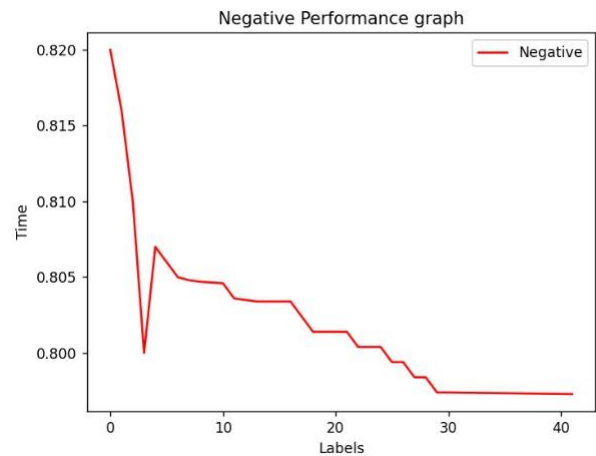

Fig 4: Negative performance Graph

The above graphs shows how the model will be processed for the Initially the negative feedback labels model building complexity is more after continuous iteration, the model to predict the negative feedback is built with less time.
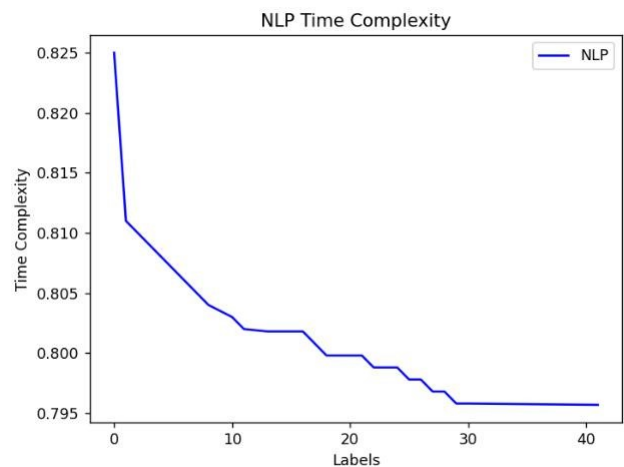

Fig.5.Time Complexity

As we have used NLP as the major algorithm to identify the feedback type the model performance should be maintained. So to build the model with high efficiency and the performance the model is built with epochs, as the epochs increase the model performance will also be increased.

V. CONCLUSION

The problem of grading-based feedback is that it does not reveal the students' true feelings. Textual feedback is utilised to address this shortcoming. Sentimental analysis allows an institution to learn about a student's feelings based on their input. They can improve the quality of instruction and the campus atmosphere as a result of this analysis. As the number of internet users grows, social media, and other factors, sentiment analysis has become a

major research topic. We discovered additional factors that had a significant impact on determining the polarity of student comments in this study. The feature impact analysis is then carried out by calculating the information gathered for each feature in the feature set and reducing the feature set using the results. To do sentimental analysis in this paper, the feedback was first gathered, then supplied to the trained model, and the sentiments were identified. The major purpose of this project is to use NLP and PoS tagging techniques to classify texts based on their sentiment. This method of extracting sentiment from text entails locating the the property of sentiment derived from the sentences Because the final product is presented graphically, the user may more easily fully understand polarity result.

REFERENCES

[1] Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., Wani, M. A. (2021). Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study. Applied Sciences, 11(9), 3986.

[2] M. Fernandez-Gavilanes, T. Alvarez-Lopez, J. Juncal-Martınez, E. Costa-Montenegro, and F. J. Gonzalez-Castano, "Unsupervised method for sentiment analysis in online texts," Expert Systems with Applica-tions, vol. 58, pp. 57–75, 2016.

[3] N. Altrabsheh, M. Cocea, and S. Fallahkhair, " ," in Adaptive and Intelligent Systems. Springer, 2014, pp. 40–49.

[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004, pp. 168–177

[5] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[6] M. Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teachers evaluation," Applied Computational Intelligence and Soft Computing, vol. 2016, p. 1, 2016

[7] M. L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analy-sis," Technical Report, 2011.

[8] 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.

[9] Tanvi Hardeniya 1, Dilipkumar A. Borikar, Dictionary Based Approach to Sentiment Analysis - A Review, International Journal of Advanced Engineering, Management and Science (IJAEMS), Volume 2, 2016

[10] Bhagyashree Gore "Sentiment Analysis of Parent Feedback" to gauge their Satisfaction Level, pp. march, 2018.

[11] Niranjan, B. S., & Hegde, V. (2021). Higher Education Enrolment Query Chatbot Using Machine Learning. In Proceedings of International Conference on Intelligent Computing, Information and Control Systems (pp. 263-274). Springer, Singapore.

[12] Hegde, V., Gnanajyothi, M., Pallavi, M. S. (2021). Perception of Students on E-Learning during Covid-19 Period and Comparison of Online and Traditional Learning. In 2021 International Conference on Intelligent Technologies, CONIT 2021.

[13] Devasia, T., Vinushree, T. P., Hegde, V. (2016, March). Prediction of stu-dents performance using Educational Data Mining. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE) (pp. 91-95). IEEE.

[14] Hegde, V., Meghana, H. N., Spandana, R., Pallavi, M. S. (2021). Predicting students performance through behavior and computational thinking in programming. In ICT Systems and Sustainability (pp. 417-429). Springer, Singapore.

[15] Hegde, V., Prageeth, P. P. (2018, January). Higher education student dropout prediction and analysis through educational data mining. In 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 694-699). IEEE.

[16] Tripathi, Milan. "Sentiment Analysis of Nepali COVID19 Tweets Using NB, SVM AND LSTM." Journal of Artificial Intelligence 3, no. 03 (2021): 151-168

[17] Pandian, A. Pasumpon. "Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis." Journal of Soft Computing Paradigm 3, no. 2: 123-134.