

# Analytical Framework for Binarized Response for Enhancing Knowledge Delivery System

Chethan G S<sup>1</sup>

Assistant Professor, Department of Information Science and Engineering, JNN College of Engineering, Shimoga, Karnataka, India

Vinay S<sup>2</sup>

Professor and Head, Department of Information Science and Engineering, PES College of Engineering, Mandya, Karnataka, India

**Abstract**—The student feedback offer effective insight into their experience of knowledge transfer, routinely collected in academic institutions. However, the existing research literature lacks reporting whether the comments in education system are helpful or non-useful. Most of the existing works are limited to sentiment polarity computation only, and teacher evaluation is carried out without considering the aspects of the teaching. This study analyzes student comments and classifies comments as applicable and non-useful for the teacher scoring system. In the proposed research, the data considered is the student feedback collected from the teacher rating website. The study performed phase-by-phase text data modeling. First, exploratory analysis is carried out on the student feedback dataset to understand text data characteristics and features. Based on the exploratory analysis, appropriate steps are determined to perform preprocessing operations for data cleaning. Using natural language processing context, the study only focuses on removing stop words and common words that belong to both useful and non-useful contexts. BoW model is considered for features extraction, and two probabilistic supervised machine learning models are used for comment classification. The study outcome exhibits that Gaussian Naïve Bayes outperforms Multinomial Naïve Bayes in accuracy, precision, recall rate, and F1-score.

**Keywords**—Education; knowledge transfer; machine learning; natural language processing; student feedback

## I. INTRODUCTION

Education is the key to success and serves as a foundation that enables people to learn about the society and world around them. People have a preconceived notion that getting an education gives them basic survival skills and is essential for good citizenship. However, education provides many opportunities in life through acquiring fundamental skills, values, ethics, morals, and life survival skills. Over the past decades, the trend of education has undergone tremendous evolutions. With the advent of technology tools and the Internet, people have raised many questions about the best teaching model. The upsurge of growth in the educational sector has a long history. The initial increase is related to personal background involving learning and memorization [1]. The upturn appeared in the education system in the early twentieth century regarding mass learning [2]. In the late 1990s, with the advent of computers and the Internet, teaching and learning methods have changed significantly. As a result, the trend of distance learning emerged parallel with the learning from books, libraries, and teaching on the blackboard. Over time, a revolutionary advancement in electronic and

network communication technologies has changed education. It induces a trend of online learning and knowledge delivery or transfer as a service at a global scale. The availability of digital platforms in the form of Information and Communication Technology (ICT) utilizes a knowledge-based learning model that will be made available to everyone by experts in various educational topics and domain disciplines [3-4]. The vision of education to all is being evidenced through different e-learning systems and MOOCS [5-6] like Khan Academy, Course era, NPTEL, etc. Various educational institutions and an educational corporate [7-8] race this. Such Institutions or learning platforms are continually looking for an efficient mechanism that can be utilized to improvise knowledge transfer and learning services. This improvement can be achieved by collecting feedback from students about the knowledge transfer by their teachers. According to the researchers [9], students' feedback provides a vital means for analyzing the student's level of understanding, attitude towards learning, and scoring teachers concerning the quality of knowledge transfer in the particular class or session. Thus, the students' feedback must be addressed efficiently by learning and education platforms to successfully enhance the quality of education in terms of helping to improve student achievement that can benefit society and the nation.

Students' feedback consists of both quantitative and qualitative information. Quantitative information such as personal feelings, beliefs, sentiments, and desires can be identified and handled easily. However, the qualitative data such as useful and non-useful comments may consist of positive and negative sentiments. Still, selecting helpful comments requires an effective classification mechanism to utilize the improvising education model and knowledge transfer services. In fact, in most cases, quantitative comments are considered, while qualitative comments are either treated manually or entirely ignored. Qualitative analysis can be performed with less data, but if the amount of information is large, a challenge will occur and be considered a text mining problem. Researches in the context of qualitative analysis of feedbacks in the education sector are relatively new. The previous study reported their contribution to student reviews mining using Natural Language Processing (NLP). The NLP-based approach uses dictionary- methods that use sentiment words match them with textual terms of comments to determine the polarity, how positive, negative, or neutral the words are like happy, sad, satisfied, dissatisfied, or angry [10-12]. Also, Machine Learning (ML) techniques such as Support

Vector Machine (SVM), clustering, neural network, and many more have been utilized in the literature to describe student understating from the various aspects level of teaching methodology, communication skills, and technical skills knowledge, etc. Though sentiment analysis plays a vital role in assessing learning effectiveness. However, most of the existing works are limited to finding the polarity of feedbacks. They do not focus on understanding and categorizing useful and non-useful review/comments patterns to effectively improve the learning organization's performance and formulate plans to enhance the knowledge transfer by teachers and students' learning experience.

Despite the variety of efforts and solutions available in the existing literature, the proposed study considers classifying helpful and un-useful comments as a text classification problem and offers an alternative approach to help score teachers, improve knowledge transfer services. In the proposed work, NLP and ML are applied to student feedback to classify valuable and useless comments benefiting learning platforms to overcome potential barriers to effective knowledge transfer services and the quality of learning processes. The idea of the proposed system is to present a novel computational framework to enhance the teaching-learning system in a digitized educational system. The importance of the research is briefed with respect to key contributions of this paper include.

- The paper highlights all the significant work carried out in a similar domain to extract unaddressed research problems.
- The paper presents a novel and simplified computational framework capable of analyzing the binary response in the form of valuable and non-useful remarks.
- The paper presents an exploratory analysis of the dataset of student response data from a teacher review site to recognize the characteristics of the feedback dataset and understate the requirement of preprocessing steps.
- The paper presents an intelligent filter to score the performance of teachers based on categorizing functional and non-useful patterns of comments using the probabilistic model-based supervised classifier.

If all the above stated research contribution is implemented than a novel framework is evolved which is capable to automating the teaching-learning experience enhancement, which doesn't exist in current research work or any commercial application. Apart from the above stated research contribution, the limitation of proposed study are viz. i) it doesn't address any distributed storage management over data center or in warehouse, ii) it doesn't address any form of security encapsulation for the extracted knowledge from the analyzed data, and iii) it doesn't consider analyzing any other language apart from English. The remaining parts of this paper are organized in the following manner: Section II. Presents brief discussion about existing techniques used for enhancing the education delivery as well as a management system. Section III discusses the motivational factors about this topic

and briefing of significant reasons behind the proposed work. Section IV describes a proposed solution to the identified research problem and detailed research methodology describing the exploratory study, preprocessing, and classification of valuable and non-useful comments. Section V discusses the experimental results presented to evaluate the proposed system's performance, whereas Section VI concludes the overall work carried out in this paper.

## II. RELATED WORK

This section discusses the related work regarding improvement in education based on student attitude analysis and teaching evaluation using feedback.

In recent years, researchers have begun to score teachers and evaluate student understanding and attitude. For instance, Beatrice Tucker [13] investigated feedback and reviews from an Australian University's aggressive or unprofessional. The investigational analysis showed that 20 comments were recognized as abusive directed towards professors and teaching, and 34 comments were recognized as unprofessional directed towards education units. A recent study by Boca [14] showed an interesting work on analyzing factors that impact student learning and attitude towards online classes during the current pandemic situation. The author surveyed by preparing a questionnaire from 300 students from technical university Romania. The findings of this study reveal that online classes and materials are beneficial for most students in a pandemic situation. But also, students found this approach to learning stressful. This study has provided good background in the context of improving education to meet the needs of new generations. Similar efforts have been made by Wu et al. [15] and Zughoul et al. [16], where the authors have analyzed students' performances and attitudes towards teachers.

The remarkable use of text mining approaches to examine student reviews and comments has led to an excess of methods. Various ML and NLP-based approaches have been applied to the education field. The work of Dhanalakshmi et al. [17] examined different ML approaches and demonstrated effective of each ML technique under consideration. The authors showed Naïve Bayes is the most suitable ML approach to compute the probability of input comments directed to specific attributes. Meanwhile, the authors in the study of Nasim et al. [18] reported lexicon-based text analysis does not require extensive data and high processing power. Qi and Liu [19] used a latent Dirichlet allocation scheme to perform texting mining over the feedback provided by the student for MOOC courses. A matrix was constructed to describe comments regarding courses. Bi-long-short-term memory (LSTM), a deep learning approach, is used to classify the sentiments from the comments.

Liu et al. [20] developed a model opinion and topic mining to forecast the worldwide acceptance level of MOOC courses. The authors have obtained sentimental comments regarding improving knowledge transfer strategies and courses. Similarly, Weng et al. [21] designed a computing model based on the feedback regarding MOOC courses. The presented model was developed based on the sentiment analysis to help education providers understand the reason for changing students' sentiments and accordingly perform course

adjustment and teaching methods. The work carried out by Koufakou et al. [22] adopted the data analytics method to mine meaningful information from student reviews to assist teachers, and educational organizations gain insights into students' emotions and attitudes. Another work done by Leong et al. [23] considered messaging texts for carrying out teaching assessments. The authors have initially classified different contexts presented within the text messages and performed data modeling using a multi-level data analytics approach. The study finding shows sentiment analysis-based model is better to conduct teacher scoring and teaching evaluation. The authors in the study of Kumar and Jain [24] introduced a model in which significant features are first analyzed.

Then, TFIDF scores were used to select features for the classification. The classification of comments is done based on the sentiment analysis-based polarity prediction using Naïve Bayes. The result indicated that 81.06% of features were correctly predicted, and 89.67 % of accuracy was achieved in the comment classification. The researchers in Chetan and Vinay's [25] study focus on optimizing data storage related to knowledge delivery systems in a cloud-based e-learning system using an advanced data analysis method using a semantic-based approach. Kastrati et al. [26] presented work on systematic mapping of existing studies on the sentiment analysis of Students' feedback with dictionary-based approaches and Learning models. The findings of this study indicated that despite many challenges, education and teacher evaluation are booming concerning the application of learning mechanisms. The authors have also highlighted the requirement of structured datasets, standardized solutions, and increased focus on emotional expression and classification. A significant study carried out by Skedsmo and Huber [27] has presented an investigational analysis based on teacher evaluation regarding the valid measures and teacher involvement. The authors discussed the various issues associated with procedures adopted in the existing studies for teacher rating in this study. The author's contribution is an essential question of the relevance between teacher rating policies and existing research procedures. Zerihun et al. [28] showed that student learning experience and achievement indicators rated teachers similarly. The authors have developed a questionnaire oriented on the students' experiences, and a hypothesis in the questionnaire is made by exploring the literature. The hypothesized evaluation parameters were designed that consist of different factors such as evaluation and review, course structure and presentation, self-assessment, and engagement.

A study conducted by Sindhu et al. [29] presented a vital work relevant to the current research. The authors have focused on the extraction of qualitative information from the student's feedback. In this work, the authors considered various aspects of teaching and used supervised learning-based sentiment analysis using multi-layer LSTM. The first layer classifies the characteristics described within the comments, and further aspect-based sentiments polarity is predicted. The dataset used in this study is prepared based on the student's observations from the specific educational institution as a use case. The study outcome indicated good

accuracy regarding aspect-based prediction with 91% and sentiment polarity with 93%. This study has provided significant concern as it focuses on different aspects needed in the teacher evaluation process. However, none of the other studies considered teaching characteristics and other necessary attributes to classify useful and non-useful features from the student feedback data. Most of the existing research is limited to polarity computation, which is suitable to highlight the emotion and sentiments of students but not much significant to score teachers. Since the student's comments and feedback often consist of irrelevant, unprofessional, abusive words, which need to be analyzed whether it is valuable or not valuable for the context of scoring or rating the teacher. The following section presents the motivation and reason behind this study.

### III. MOTIVATION

The discussion in the first section is carried out from the perspective of the education system and its improvement. In the present section, the discussion is carried out from the teachers' perspective and the influential factors that motivated and the reason behind the proposed work. A teacher has a significant role in everyone's life. Teachers engage students' minds to help in gaining knowledge and understanding the subject of interest. The teaching aspects from different perspectives are illustrated in Fig. 1.

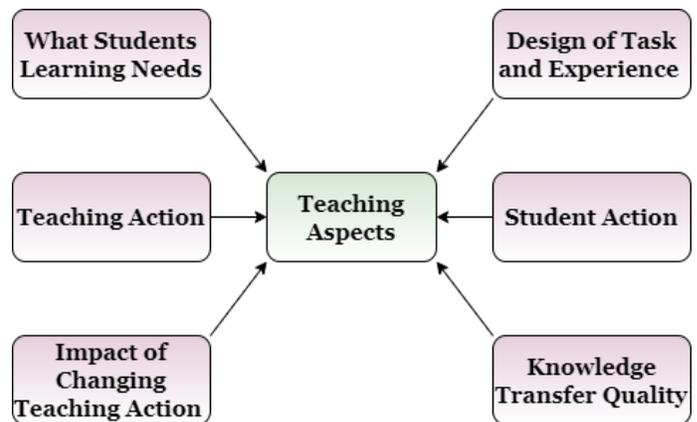


Fig. 1. An Illustration of Teaching Aspects.

In education, teachers have to contribute in several ways, including understanding their students, what student learning needs, what type of task design and assessments should be included. Another important thing is determining the quality of knowledge or teaching action required to create positive associations with students, make an operative learning environment, and establish proficiency and accountability. However, society is constantly evolving, and the performance requirements of education organizations and online learning platforms are also changing. The teaching profession is no exception except to bring about a change in academic life and professional commitments. In this regard, the requirements for teachers have become more and more complex. Although teachers try to do their best, sometimes the quality of knowledge deliverance is questioned. Reflecting on these changes and requirements in the current scenario, teacher scoring or rating systems have become more challenging and

complex. Rating teacher refers to accessing teaching skills, the approach of knowledge transfer, and teachers' achievement. Based on the discussion in the previous section, the sentiments analysis from students' feedback provides correlational attributes to score teachers and courses. Particularly in the context of education, the positive sentiments are believed to have good learning from the teachers, while negative sentiments believed to be knowledge transfer by teachers for their classes is not satisfactory. The sentiment analysis from textual data (student's feedback or comments) raises many technical challenges. A key challenge is understanding the essential features and classifying helpful and non-useful comments regarding scoring teachers. However, none of the existing research works have focused on this challenge. The proposed study believes that analyzing the sentiments from textual data is highly contextual as the wording meaning varies across different domains. For instance, in terms of education context, the word "extremely" indicates a negative sentiment in the comment-1 "She moves extremely fast, teaches different methods than the book."

However, the same word 'extremely' indicates a positive one in comment-2 "Extremely an important class". The word 'extremely' in both sentences reflects different sentiments. Let us consider another example with the same word that reflects positive sentiment in comment-3 "This is her favorite class and is extremely passionate about it". The comment-3 reflects positive sentiments, but it is not helpful because it does not describe any clear information and basis for improvement. Careful observation reveals that though comment-1 and comment-2 reflect different sentiments, they can be considered useful comments. They have significant intentions that contribute to improvising teaching methods and an effective learning environment. To get more clarity, let us see another example in terms of a different domain. The word "quickly" indicates a negative sentiment in the comment, "The explanation ended too quickly!". However, in a restaurant context, it means a positive sentiment in the comment "The service provided quickly". These examples show how important it is to understand the critical feature of comments to determine the valuable and non-useful comments regarding educational context.

It should also be noted here that useful and useless comments can be composed of positive and negative emotions, and the task is to determine or classify useful and useless. Unfortunately, most existing literature does not consider the above factors and generally divides comments into positive, negative, and sometimes even neutral. They do not consider the correlation between sentiment and the usefulness of comments, whether it is understood as a sentiment classification problem without considering salient features and contextual significance. In turn, this raises serious questions about the current approach to the rating or assessment of teaching aspects. In this sense, the vagueness in the existing literature leaves education providers and learning organizations with a constrained direction about an effective way to make new policies towards improving courses, knowledge transfer services, and reforming other units of education. The proposed research aims to present an effective intelligent filter to score teachers based on identifying helpful

and non-useful comments from student feedback. This work can be defined as a problem of text classification from the given comments in textual representation concerning helpful comment (Uc) and non-useful comment (Nc) aspect:  $\{Uc, Nc; \exists \text{ comments } (c)\} \leftarrow \text{function } f \text{ such that } (c) = \{Uc, Nc\}$ . The study performs phase-by-phase analytics of textual data to determine the main aspects to be taken into account for designing a teacher scoring model.

#### IV. PROPOSED SYSTEM

This section discusses the methodology and implementation strategy for the proposed system. The suggested work presents a significant contribution in the field of learning analytics. The study has been carried out to analyze substantial comments from the students' feedback rather than the sentiment polarity computation. The study considers that if good data analytics is carried out, then helpful and non-useful comments can be distinguished efficiently, positively impacting the teacher scoring system, thereby improving the quality of knowledge transfer in e-learning systems and academic institutions.

The workflow of the proposed system architecture is shown in Fig. 2, which has five core components: the first collection of text from the student's feedback; second exploratory analysis to understand characteristics of data and steps required in data cleaning in the third preprocessing, where irrelevant information associated with text data is excluded, and unique words specific to helpful comments are analyzed; fourth feature extraction is carried out using bag of word (BoW) model making text data suitable to fit in classification model; then designing an intelligent filter for classification of a useful and non-useful comment using the supervised classifier. All these processes are performed phase-by-phase, and a text cloud is used to visualize the essential words related to helpful and non-useful comments. The system uses the open-source Python programming language to perform modeling and implementation of the entire system.

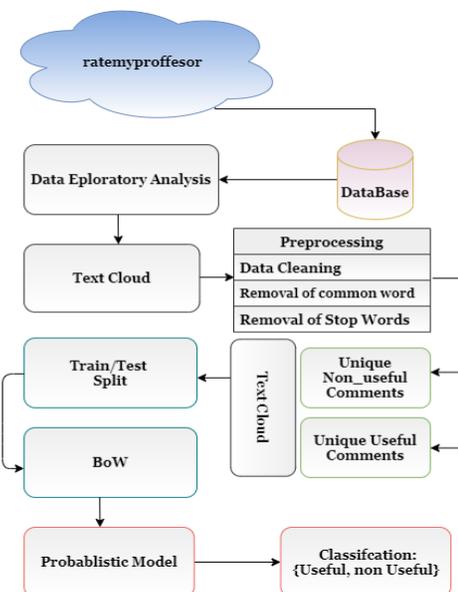


Fig. 2. Systematic Workflow of the Proposed System.

A. Data Aggregation

The data aggregation in the proposed research work is student feedback from the teacher review site. Academic institutions and online learning platforms often do not disclose their internal input or review by students to assess teaching quality. Several online platforms have emerged, which allow students to evaluate their teachers publicly. The most popular online evaluation platform is RateMyProfessors.com. Fig. 3 illustrates the online student learning and feedback sharing over the learning management system.

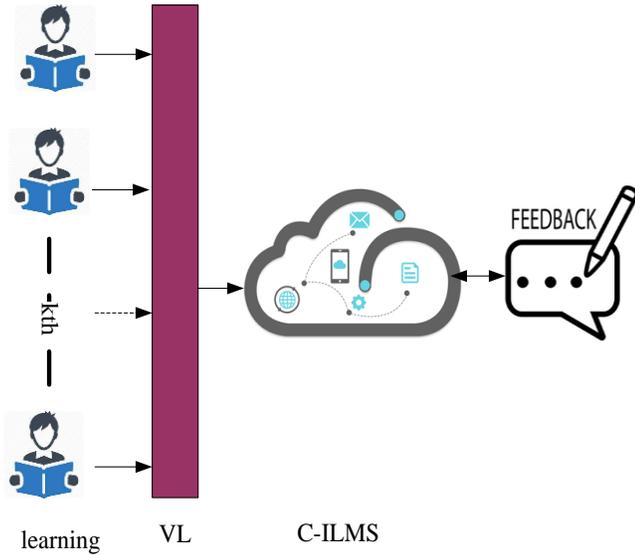


Fig. 3. The Conceptual Scenario of Learning and Feedback Sharing.

In Fig. 3, the conceptual scenario of e-learning or online system and the feedback sharing process is demonstrated. The students get education and knowledge via a virtual layer (VL) connected with a Cloud-Enabled Intelligent Learning Management System (C-ILMS). This kind of platform also allows students to evaluate their teachers anonymously. Students present their opinion through review or comment from different aspects. For teacher rating, clarity, helpfulness, and difficulty are considered the most. While rating the courses and learning platform, students give comments considering different aspects such as academic institution reputation, education-related services offered, test and assessments, activities, and opportunities. "RateMyProfessors" (RMP) is a teacher review website where students can provide an open comment and express their emotions very transparently. RMC provides a simple way of sharing what students experience in a course and their professor's ratings. On this website, "ratings" refers to students' responses to a single survey question about overall professor quality that includes rating indicators, content knowledge, interest, clarity of lectures, etc.

In the proposed study, data of student feedback regarding teacher ratings are collected from the RMP website. The collected dataset considers comments indicating a review of several professors. A web scraping technique is used on RMP in the data collection process. The dataset represents only textual and numerical data, therefore of small dimensions.

B. Data Exploratory Analysis

In the system model, the data intuition is first created utilizing the word cloud. The entire dataset in the unstructured format is considered as a universal set ( $\Omega$ ) which consists of subsets of helpful comments ( $U_c$ ) and non-useful comments ( $N_c$ ) such as  $U_c \subset \Omega$  and  $N_c \subset \Omega$ . Table I presents the sample view of the student feedback dataset.

Table I shows the data reading process from the dataset provided by ratemyprofessor.com. The first step in this data analysis phase is to observe the most common word in the comments or text specific to users and no useful comments. The dataset is labeled as class usefulness with two possible states {usefulness-state ( $U_c$ ):1, non-usefulness ( $N_c$ ):0}. The total number of samples ( $T_s$ ) in the dataset is 562, whereas only 358 samples are applicable and 204  $\in$  non-useful states. Therefore, the proportion of usefulness-state and non-usefulness state can be found using equations 1, 2 and 3, which are  $x=63.7\%$  and  $y=36.3\%$ , respectively.

$$T_s = \sum(U_c, N_c) \tag{1}$$

$$x = \frac{100 \times U_c}{T_s} \tag{2}$$

$$y = \frac{100 \times N_c}{T_s} \tag{3}$$

To understand the distinct pattern between  $U_c$  and  $N_c$ , the mean length of the feedback comment text is computed using equations 4 and 5 for both  $U_c$  and  $N_c$ .

$$\mu_{U_c} = \frac{1}{n(U_c)} \sum_{i=1}^{n(U_c)} \sum_{j=1}^N U_{u_i}(w_j) \tag{4}$$

$$\mu_{N_c} = \frac{1}{n(N_c)} \sum_{i=1}^{n(N_c)} \sum_{j=1}^N N_c(w_j) \tag{5}$$

The mean ( $\mu$ ) length of each class of the comments computed is shown in Fig. 4. Computation of the mean value will offer a simplified way to understand the type of comment for further assisting in data analysis in upcoming operation steps. On an analysis as in Fig. 1, it is found that the average size of the  $U_c$  is 46.26 words, and the average length of the  $N_c$  is 26.59 words per comment. This numerical score offers a clear picture for the filtered formed of statements concerning the usefulness and non-usefulness state of the response.

The primary classification is possible even based on the word count. However, it may not be so in a real-time scenario, as sometimes or many times the average length of expression might be interchangeable. The study also presents text cloud to get more insight into both  $U_c$  (Fig. 5(a)) and  $N_c$  (Fig. 5(b)) as follows.

TABLE I. SAMPLE VIEW OF COMMENTS IN THE DATASET

Index	Comments	Label
0	He was boring and ruined psychology for me. No	Nc
1	challenging course. I thought I was go	Uc
2	Hard to understand at times. Tests are ok. Not.	Uc
3	SO glad to be the hell out of this class	Nc
4	great teacher and very smart, the class is very	Uc

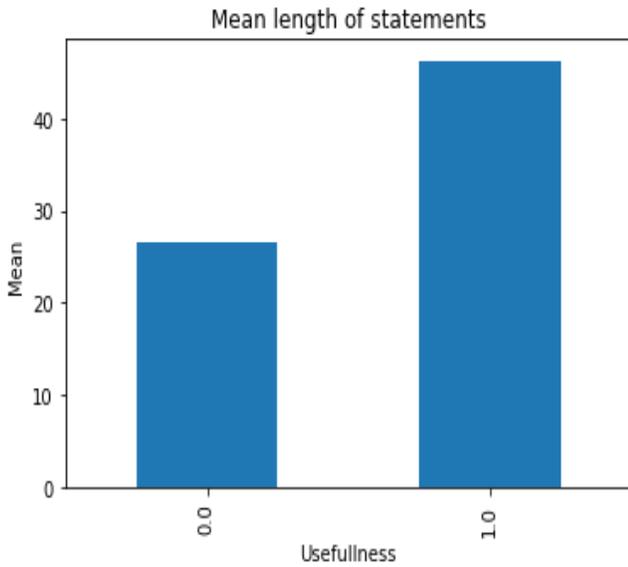


Fig. 4. Mean Length of Comment Text for Both  $U_c$  and  $N_c$ .

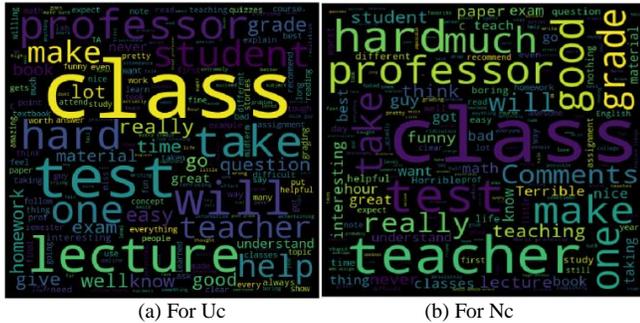


Fig. 5. Visualization of Text Cloud.

The study then checks for the most common words in the useful and non-useful categories. If the comments are insightful, it is considered useful, but the not constructive comments are not useful. The mapping of {useful, non-useful} → {1,0}. The short comments are also ignored. The text cloud is constructed with useful and non-useful. However, if the size of the word in the text cloud is big and bold then repeated, then it is more repeated, significant and smaller means less repeated.

The analysis obtained from text cloud Fig. 5(a) and Fig. 5(b) provides essential information that will help make the dataset suitable for the input to the model input by normalizing and discarding irrelevant and repetitive data. This process positively impacts the computational dependency of the model, i.e., it lessens the computational requirement in the feature representation and learning process. In this regard, the proposed exploratory analysis further explores the unique words in the  $U_c$  and  $N_c$ . A closer look indicates that many words stop words, so the next step is to remove the common and stop words. Fig. 6 highlights some stops words  $\in U_c, N_c$ .

i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', 'you're', 'you've", "you'll""you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven'

Fig. 6. Illustration of Some Stop Words.

The Stop words are the words that contain less information and are repetitive in the same line of comments as well as other lines of comments that  $\in \{U_c, N_c\}$ . The amount of information on a particular word is measured by entropy as described in equation 6 and equation 7 as follows:

$$H = -\sum p(x). \log(p(x)) \quad (6)$$

Where  $H$  denotes entropy of information reflected from particular words,  $p(x)$  represents the probability of the word.

$$p(x) = \frac{\text{No.of time the word has been repeated}}{\text{Total number of words}} \quad (7)$$

If a word is most probable, it is closer to 1, and  $\log(1)$  is zero, meaning it carries no information. However, it can be seen that a negative sign at the beginning of equation (6). Hence, the less repeated word carries more information since the value of  $\log(p(x))$  is more. However, if a word is not appearing at all, then also it carries no information. In order to represent this, the  $p(x)$  is multiplied to  $\log(p(x))$ .

Furthermore, when considering valuable comments, the punctuation cannot be ignored in the proposed study. People writing serious comments or helpful comments will not use punctuation marks like '!' or '?' or multiple dots. Hence, the words 'completed' and 'completed!!' give a very different meaning in the proposed study. The first word reflects the writer's seriousness, whereas the second word represents either excitement or frustration of the writer. When a professor or a teacher is being rated, such comments, which reflect emotion more than real criticism or appreciation, must be avoided. Therefore, the proposed study considers the only removal of stop words using preprocessing algorithm.

### C. Data Preprocessing

The proposed system initiates a preprocessing operation to ensure better retention of data. Stop words commonly exist in all forms of text, and hence they should be first removed. Therefore, the preprocessing algorithm is designed to remove the stops words from the entire text corpus in this implementation phase.

The description of the algorithm is as follows:

**1. Algorithm for Preprocessing**

```

Input: Dataset (DS)
Output:Preprocessed feedback_texts
Start:
Load: DS → f1('DS')
        function: rmv_stop(text)
        for each non – useful comments df do
check: text(class == 0) do
W.join(rmv_stop('text') do
        join: non – useful → W
        for each useful comments df do
check: text(class == 1) do
W.join(rmv_stop('text') do
        join: useful → W
        Apply→text.split ( Words ∈ {Uc, Nc} )
        List useful comments & non useful comments
End
    
```

The algorithm has an inclusion of multiple parameters which are briefed as follow: i) *DS* is used as a dataset acting as input to proposed system, ii) *df* is used as non-useful comments which is required to be filtered out, iii) *W* is used as a useful comments which is considered for further analysis, iv) *U<sub>c</sub>* and *N<sub>c</sub>* is final matrix which restores useful and non-useful comments respectively. The algorithm takes all the textual feedback data and applies the *rmv\_stop* function to remove stop words without scarifying the meaning of the sentences. The explicit function *rmv\_stop* is used from the python library, representing all the stops words defined in the English language. Each sentence is passed into the functions, and all the stops' words are removed.

For each *N<sub>c</sub>* in the text field of the dataset, a condition is evaluated that class of text is found to be 0 then remove stop word in the non-useful comments and joined to word vector *W*. Another condition gets evaluated to remove stop words from the valuable comments and further filtered helpful comments were appended to the *W*. Another function is then used, i.e., text. Split to separate the sequence of text and lists the word belonging to the *N<sub>c</sub>* and *U<sub>c</sub>*. Fig. 7 indicates the word cloud. The study explores to get insight on the common word contained in both *U<sub>c</sub>* and *N<sub>c</sub>*. Table II highlights common words extracted during analysis. A closer look into Table III and analysis of the word cloud shows some common words in *U<sub>c</sub>* and *N<sub>c</sub>*. Therefore, looking at the Venn diagram, only valuable words are considered, and common words are avoided.

Fig. 7 Venn diagram provides an analysis that words belong to both *U<sub>c</sub>* and *N<sub>c</sub>* such that:  $U_c \cap N_c$  words are avoided. The familiar words are non-constructive, and only non-useful words are filtered. Those words will be considered if the phrase is uncommon in the helpful comment, and all the common words will be avoided. Similarly, uncommon words are selected for *N<sub>c</sub>*, and common words are avoided. A frequency distribution plot is mentioned in Fig. 9 and Fig. 10 for unique words  $\in \{U_c, N_c\}$ .

The exploratory data analysis provided an understanding of data towards a better decision in the preprocessing, where only stops words and common words are removed. The study does not carry out any removal of punctuation as it has significant intent in the comments. The above processes provided a substantial analysis of the data, which positively impacted the feature reduction and training of the model.

From the visuals of Fig. 8, it is clear that the proposed system performs filtration of the words on the basis, thereby preprocessing the text.

TABLE II. SAMPLE OF COMMON WORDS

Rounds	Extracted Common Words
1	'extremely', 'difficult', 'course.', 'T', 'thought', 'T', 'was'
2	'going', 'to', 'have', 'to', 'retake', 'it', 'many', 'have', 'The'
3	'labs', 'are', 'hard', 'and', 'quite', 'time', 'and', 'the', 'pop'
4	'quizzes', 'are', 'T', 'think', 'the', 'tests', 'should', 'have'
5	'questions', 'but', 'be', 'If', 'you', '2', 'multiple', 'you'
6	'drop', 'a', 'grade.', 'this', 'course', 'is', 'VERY'
7	'Hard', 'to', 'understand', 'at', 'times.', 'Tests', 'are'
8	'Not', 'too', 'with', 'great', 'teacher', 'and', 'very'
9	'smart', ',', 'the', 'class', 'is'

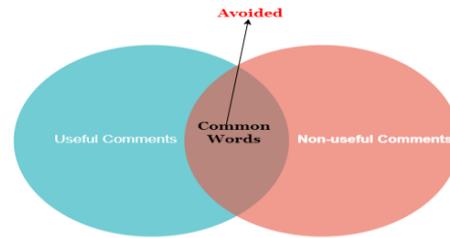


Fig. 7. Intersection of the *U<sub>c</sub>* and *N<sub>c</sub>* words is avoided



(a) Text cloud for *U<sub>c</sub>* after removal of stops, words.



(b)Text cloud for *N<sub>c</sub>* after removal of stop words.

Fig. 8. Text cloud analysis after removal of stop words.

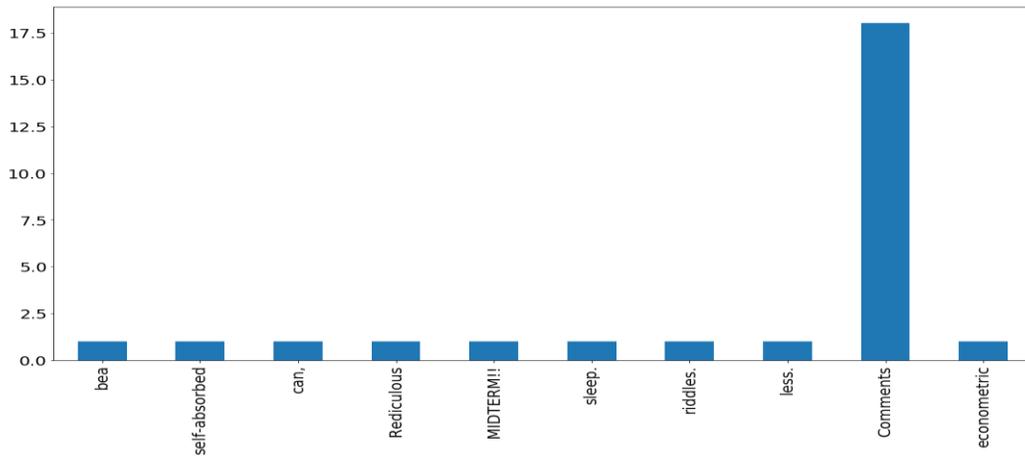


Fig. 9. Frequency Distribution of Words ∈ N<sub>c</sub>.

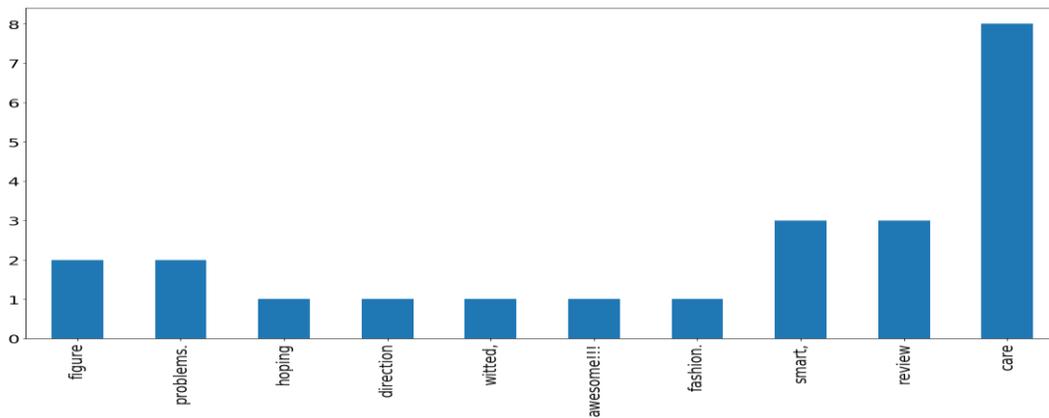


Fig. 10. Frequency Distribution of Words ∈ U<sub>c</sub>.

The entire process discussed above is carried out phase-by-phase, providing effective feedback data modeling and preprocessing steps. The next step discusses the classification module, where the dataset split process is carried first, and different features are extracted using the BoW model. The algorithm for U<sub>c</sub> and N<sub>c</sub> classification is discussed as follows:

## 2. Algorithm for Classification

---

**Input:** Dataset DS  
**Output:** Classified comment class (C1, C2)  
**Start:**  
**Init DF** → attri(DS)  
 For each text ∈ **DF**, **do**  
**Preprocessing:** Use **Algorithm 1**  
     Call **function:** rmv\_stop(text)  
 Split DS → [Tr, Ts]  
 Extract Feature ← [Tr]  
     Feat → []  
     BoW → f2(Tr)  
     **Feat** → Feat.append.Bow  
 Model Training → [Multinomial NB, Gaussian NB] **do**  
     **Training** → model.train(Feat, Tr)  
     Model Deployment: **do**  
     **Testing** → model.test(Feat, Ts)  
 Testing done: **Evaluate** [Accuracy, F1 Score, Recall, Precision]  
**End**

---

The algorithm takes the input dataset. After processing, it provides a classified class of comments. Initially, it loads the attributes of the dataset as a data frame. In the next step, the algorithm for the calls pre-processing operation is described in the first algorithm. In the next step, dataset split operation is carried out in terms of 80:20 ratio where 80% of the dataset is subjected to training set and 20% dataset as testing set.

Further feature extraction and vectorization are performed to make the training dataset a suitable input for the classification. The study uses the BoW model to extract features followed by the stemming and lemmatization process. In the proposed research, the probabilistic model, i.e., the Naïve Bayes approach, is used to perform classification as it is free from extensive training data and capable of handling both discrete and continuous data. The Naïve Bayes is based on the method of Bayes theorem numerically expressed in equation 6 as follows:

$$P(x|y) = \frac{P(x|y)P(x)}{P(x)} \quad (8)$$

In above equation(8), where (x) denotes the prior probability being true, P(x) refers to the probability of the data, P(x|y) is the probability of presumption x for the available data y, and P(x|y) denotes the probability of y given that presumption x is true. In the proposed work the study implements two types of Naïve Bayes algorithm such as

i) Multinomial Naïve bayes, and ii) Gaussian Naïve bayes for performance analysis. The model is trained in a supervised manner using the function model. The train takes input argument of feature and training set. The performance evaluation model is carried out with multiple performance metrics.

#### D. Performance Metrics

This section discusses the performance metrics considered in the proposed system for model evaluation. Accuracy is the ratio of the number of correct predictions over a total number of predictions; it can be computed as follows:

$$\text{Accuracy} = \frac{\alpha + \alpha'}{\alpha + \alpha' + \gamma + \gamma'} \quad (9)$$

In above equation(9), the variable  $\alpha$  represents true positive,  $\alpha'$  is the true negative metric,  $\gamma$  indicates true negative and  $\gamma'$  represents false negative evaluation metric.

The precision metric is the ratio of correct predictions over a total number of predictions made to the current class. The numerical expression for precision is mentioned as shown in equation (10).

$$\text{Precision} = \frac{\alpha}{\alpha + \gamma} \quad (10)$$

Recall rate is the ratio of correctly predicted values over the number of values in that class. Recall rate can be numerically expressed as shown in equation (11):

$$\text{Recall rate} = \frac{\alpha}{\alpha + \gamma'} \quad (11)$$

F1 score is the weighted mean of precision and recall, which genuinely represents the system's performance. It can be computed as shown in equation (12):

$$\text{F1}_{\text{Score}} = 2 \left( \frac{\left( \frac{\alpha}{\alpha + \gamma} \right) \times \left( \frac{\alpha}{\alpha + \gamma'} \right)}{\left( \frac{\alpha}{\alpha + \gamma} \right) + \left( \frac{\alpha}{\alpha + \gamma'} \right)} \right) \quad (12)$$

### V. RESULT ANALYSIS

In this section, the result and performance analysis of the proposed system is discussed. The proposed method for teacher rating based on valuable and non-useful comments is carried out in Python's numerical computing environment. Based on the probabilistic model, two supervised classifiers such as i) Multinomial Naïve Bayes and ii) Gaussian Naïve Bayes, are considered to classify helpful and non-useful comments. These classifiers are selected because they can handle large text datasets and are suitable for natural language processing tasks. Numerical outcomes are shown in Table III. The model training is carried out by 80 % of dataset text samples, i.e., 450 text samples out of 562 text samples, and model testing is carried out with 20% of text samples, i.e., 112 text samples. The performance of the proposed model is measured in terms of multiple performance metrics such as recall, F1-score, and accuracy rate. Fig. 10 shows the performance analysis of the proposed model based on the outcome achieved in terms of accuracy (%).

Fig. 11 compares the outcome achieved from the Gaussian probabilistic model and the multinomial probabilistic model. The graph trend exhibits that the Gaussian Naïve Bayes

outperforms the multinomial naïve byes classifier. The gaussian NB has achieved an 83.74% accuracy rate in the text class prediction, i.e., approximately 75 text samples were correctly predicted out of 112 text samples in the testing dataset. In the case of multi-nominal NB, only 71.15 % of the accuracy rate is achieved with the same training dataset. Total 63 text samples were correctly predicted as useful and non-useful by multi-nominal NB from the total 112 text samples in the testing dataset. The overall analysis suggests that the Gaussian Naïve Bayes classifier achieved a similar performance pattern concerning all metrics such as precision, recall, F1-score, and accuracy. Therefore, the Gaussian Naïve Bayes algorithm is most stable and suitable for comment classification. Apart from this, the proposed system is also assessed concerning processing time. Upon execution over a similar testbed, it is found that the Gaussian Naïve Bayes approach consumes less processing time than the Multinomial approach. The former method is appropriate for addressing prediction problems with multiple classes and hence they are ideal for the proposed study over the educational domain. From the numerical outcomes obtained in proposed analysis, it is seen that there are various rationale behind the improved outcomes of Gaussian Naïve Bayes approach in Machine learning in comparison to Multinomial Naïve Bayes approach. It is undeniable to state that Multinomial Naïve Bayes is one of the effective analytical approach used in natural language processing using probability concept; however, when this is exposed to larger set of dynamic data, irrespective of any domain of data, the prediction accuracy is usually lower in contrast to other schemes. Multinomial Naïve Bayes also computes the probability of all the tags in sample corpus with highest probability however, that is never satisfactory. Apart from this, it cannot be used for any regression and is restricted to forecast numerical value prediction. However, Gaussian Naïve Bayes, when exposed to proposed dataset, is witnessed to work in speedy manner and consumes less time. It is also found suitable for solving prediction of multi-class of data.

TABLE III. NUMERICAL OUTCOME

	Multinomial	Gaussian
Precision	0.726531	0.839381
Recall	0.771429	0.839286
F1-Score	0.715584	0.833227
Accuracy	0.715556	0.839286

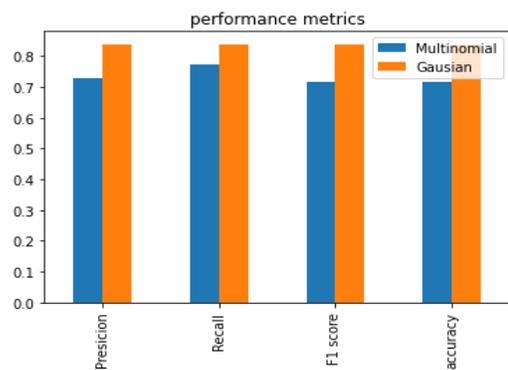


Fig. 11. Comparative Analysis.

## VI. CONCLUSION

The study aims to understand feedback data from the teaching aspect, recognize its critical feature to classify comments as applicable and non-useful to score teachers, and lead to high student achievement. The proposed model performs exploratory analysis of the textual feedbacks and identifies unique and common words concerning contextual meaning. Further, the necessary data cleaning process is carried out to make the dataset suitable for comment classification two using the probabilistic model of the supervised learning approach. With a practical and phase-wise data modeling, the proposed model that does not suffer from high dimensional feature space provides a compelling analysis of student feedback that helps improve learning and teaching aspects in knowledge transfer as service platforms, like service MOOCs and e-learning education systems. The result indicated the effectiveness of the Gaussian Naïve Bayes classifier with a higher accuracy rate than the multi-nominal Naïve Bayes classifier.

### REFERENCES

- [1] MM. Ghonge, R. Bag, A. Singh. Indian Education: Ancient, Medieval and Modern. In Education at the Intersection of Globalization and Technology 2020 Oct 27. IntechOpen.
- [2] J. Zinkina, A. Korotayev, A. Andreev. Mass primary education in the nineteenth century. In Globalistics and Globalization Studies: Global Transformations and Global Future, pp. 63-70, 2016.
- [3] A. A. Genlott, A. Grönlund, & O. Viberg, Disseminating digital innovation in school – leading second-order educational change. *Educ Inf Technol*, vol.24, pp.3021–3039, 2019.
- [4] J. Fraillon, J. Ainley., W. Schulz, T. Friedman, E. Gebhardt, Students' Use of and Engagement with ICT at Home and School. In: *Preparing for Life in a Digital Age*. Springer, Cham, 2014.
- [5] D. E. Fatumo and S. Ngwenya, "Online learning platforms and their roles in influencing pass rate in rural communities of South Africa: Massive Open Online Courses(MOOCs)," 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 2020, pp. 1-8, doi: 10.1109/IMITEC50163.2020.9334135.
- [6] A. Vaibhav and P. Gupta, "Gamification of MOOCs for increasing user engagement," 2014 IEEE International Conference on MOOC, Innovation, and Technology in Education (MITE), 2014, pp. 290-295.
- [7] K. Ingolfsdottir, "Impact of MOOCs and other forms of online education [point of view]," *Proceedings of the IEEE*, 16;102(11), pp.1639-43, 2014.
- [8] M. Alshehri, Alamri, A., Cristea, A.I et al. Towards Designing Profitable Courses: Predicting Student Purchasing Behaviour in MOOCs", Springer-International Journal of Artificial Intelligence in Education, vol.31, pp.215-233, 2021.
- [9] Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C. et al. Impact of students evaluation of teaching: a text analysis of the teachers qualities by gender. *Int J Educ Technol High Educ* 17, 49 (2020).
- [10] Sangeetha, K., Prabha, D. Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. *J Ambient Intell Human*.
- [11] Rajput Q, Haider S, Ghani S. Lexicon-based sentiment analysis of teachers' evaluation. *Applied computational intelligence and soft computing*. 2016 Oct 1;2016.
- [12] Nasim, Zarmeen & Rajput, Quratulain & Haider, Sajjad. (2017). Sentiment analysis of student feedback using machine learning and lexicon based approaches. 1-6. 10.1109/ICRIIS.2017.8002475.
- [13] Tucker B. Student evaluation surveys: anonymous comments that offend or are unprofessional. *Higher Education*. 2014 Sep 1;68(3):347-58.
- [14] Boca GD. Factors Influencing Students' Behavior and Attitude towards Online Education during COVID-19. *Sustainability*. 2021 Jan;13(13):7469.
- [15] Wu XM, Dixon HR, Zhang LJ. Sustainable Development of Students' Learning Capabilities: The Case of University Students' Attitudes towards Teachers, Peers, and Themselves as Oral Feedback Sources in Learning English. *Sustainability*. 2021 Jan;13(9):5211.
- [16] Zughoul O, Momani F, Almasri OH, Zaidan AA, Zaidan BB, Alsalem MA, Albahri OS, Albahri AS, Hashim M. Comprehensive insights into the criteria of student performance in various educational domains. *IEEE Access*. 2018 Nov 14;6:73245-64.
- [17] Dhanalakshmi, V., Bino, D., & Saravanan, A. M. (2016). Opinion mining from student feedback data using supervised learning algorithms. In 2016 3rd MEC international conference on big data and smart city (ICBDSC) (pp. 1–5).
- [18] Nasim, Z., Rajput, Q., & Haider, S. (2017). Sentiment analysis of student feedback using machine learning and lexicon based approaches. In 2017 international conference on research and innovation in information systems (ICRIIS) (pp. 1–6).
- [19] Qi C, Liu S. Evaluating Online Courses via Reviews Mining. *IEEE Access*. 2021 Feb 24;9:35439-51.
- [20] Z. Liu, W. Zhang, J. Sun, H. N. H. Cheng, X. Peng, and S. Liu, "Emotion and associated topic detection for course comments in a MOOC platform," in Proc. Int. Conf. Educ. Innov. through Technol. (EITT), Sep. 2016, pp. 15–19.
- [21] J. Weng, W. Gan, G. Ding, Z. Tian, Y. Gao, and J. Qiu, "SESM: Emotional, social semantic and time series analysis of learners' comments," in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Oct. 2020, pp. 20–28.
- [22] A. Koufakou, J. Gosselin, and D. Guo, "Using data mining to extract knowledge from student evaluation comments in undergraduate courses," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2016, pp. 3138–3142.
- [23] C. K. Leong, Y. H. Lee, and W. K. Mak, "Mining sentiments in SMS texts for teaching evaluation," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2584–2589, 2012.
- [24] A. Kumar and R. Jain, "Sentiment analysis and Feedback Evaluation," 2015 IEEE 3rd International Conference on MOOCs, Innovation, and Technology in Education (MITE), pp. 433–436, 2015.
- [25] Chethan G.S., Vinay S. (2019) Virtual Map-Based Approach to Optimize Storage and Perform Analytical Operation on Educational Big Data. In: Sridhar V., Padma M., Rao K. (eds) *Emerging Research in Electronics, Computer Science and Technology*. Lecture Notes in Electrical Engineering, vol 545. Springer, Singapore.
- [26] Kastrati, Z., Dalipi, F., Imran, A.S., Pireva Nuci, K., and Wani, M.A., 2021. Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study. *Applied Sciences*, 11(9), p.3986.
- [27] Skedsmo G, Huber SG. Teacher evaluation: the need for valid measures and increased teacher involvement. *Educational Assessment, Evaluation, and Accountability*. 2018 Feb;30(1):1-5.
- [28] Zerihun Z, Beishuizen J, Van Os W. Student learning experience as an indicator of teaching quality. *Educational Assessment, Evaluation, and Accountability*. 2012 May;24(2):99-111.
- [29] Sindhu I, Daudpota SM, Badar K, Bakhtyar M, Baber J, Nurunnabi M. Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access*. 2019 Jul 15;7:108729-41.