Journal of Engineering Education Transformations ,
Volume 38 , No. 2 , October 2024 , ISSN 2349-2473, eISSN 2394-1707

10.16920/jeet/2024/v38i2/24187

A Quantitative Approach for Appraising Quality of Online Education

R. Kaur[1], R. Sharma[2], A. A. Jha[4], V. Gaur[3]
[1]Department of Electronics, Acharya Narendra Dev College, University of Delhi, New Delhi-110019, India.
[2,3,4] Department of Computer Science, Acharya Narendra Dev College, University of Delhi, New Delhi-110019, India.
[1]ravneetkaur@andc.du.ac.in
[2]rajatsmsc23@cs.du.ac.in
[3]adi4ant@gmail.com
[4]vibhagaur@andc.du.ac.in

Abstract : Conventional approaches fall short of managing the current scale of data, underscoring the significance of big data analytics in the field of data science. The study highlights the importance of big data analytics in data science, focusing on sentiment analysis and developing a machine learning framework for identifying sentiments in product reviews and assessing Massive Open Online Courses (MOOCs). Online education proved its value to the academic community during the COVID-19 pandemic, when much of the world was at a standstill. Since the coronavirus outbreak, Massive Open Online Courses (MOOCs) have taken over many public schools and undergraduate degree programs as the new norm. Nowadays, online courses are being incorporated into traditional educational programs. Therefore, it has become crucial to develop a framework that both government and non-government organizations could use to assess public opinion before implementing online education programs. This article offers a methodology for utilizing deep learning and natural language processing techniques to examine public sentiments toward online education and courses such as MOOCs.

Public reviews from Coursera and Udemy are utilized to assess the quality of online courses. Using sentiment analysis, the proposed methodology employs the Bidirectional Encoder Representations from Transformers (BERT) model to classify public reviews into positive and negative categories. Positive and negative reviews are clustered using the k-means algorithm to evaluate the quality of MOOCs.

The study's findings indicate that MOOCs with qualified professors are beneficial for learning basic concepts. But they may lack advanced knowledge, real-world experience, and enough examples. These courses are lengthy, and students request more questions after lessons to comprehend the concept. The proposed framework assists in assessing public opinion before implementing online education and may provide valuable insight to the developers of the courses.

V. Gaur
Department of Computer Science,Acharya Narendra Dev College, University of Delhi, New Delhi-110019, India.
vibhagaur@andc.du.ac.in

Keywords : MOOC, BERT, K-means Algorithm, Clusters, Word Cloud.

1. Introduction

Electronic connectivity has become indispensable in modern society, where the use of computers, tablets, or smartphones is essential for everyday functioning. In developed countries, digital technology has already replaced manual tasks and technological advances continue to change many aspects of life. Everyone in today's interconnected world must be able to use various forms of cutting-edge technology effectively. The internet and other technological advancements have impacted many facets of society, including academia, medicine, business, banking, economics, agriculture, and other related fields. The ability to easily share and access information from anywhere in the world is a big part of what makes the Internet so appealing to people from all walks of life. This is the most effective method for fostering relationships between customers, suppliers, and businesses.

As the Internet user base grows, so does the user base of digital devices used for online shopping, banking, education, and other purposes. Due to the widespread adoption of digital technologies across a wide range of industries, data has proliferated, making it increasingly difficult to manage. In both the public and private sectors, the ability to analyse and mine data from large datasets is becoming more and more important. Nonetheless, the difficulties associated with Big Data analytics increase proportionally with data volume.

As social media sites like Facebook, Twitter, and YouTube become more popular, a lot more user data is growing. Traditional technologies are insufficient for managing the colossal amount of data being created. As a result, big data analytics has become a crucial area in data science, with an increasing number of companies developing products that employ it for market research, product development, and various other purposes.

The data science toolkit includes two key components: big data analytics and deep learning. Tech behemoths like Google and Microsoft impact current and future developments by analysing vast volumes of data for corporate research and decision-making. Through a hierarchical learning process, deep learning approaches extract sophisticated and high-level abstractions as data representations. At every level of the hierarchy, more complex abstractions are built on previously learned simpler abstractions. Deep Learning's ability to evaluate and learn from massive amounts of unsupervised data makes it an invaluable tool for Big Data Analytics, where much of the raw data is yet to be classified or labelled. Recent progress in deep learning has significantly enhanced the text-analysis capabilities of algorithms. Using innovative AI techniques in distinct ways can be a powerful means for conducting comprehensive research.

In order to determine and extract opinions and attitudes from the text, sentiment analysis is performed. Customer satisfaction is critical to the growth and success of any business, regardless of industry. Getting feedback from customers about how they feel about a company's brand, product or service is valuable. Sentiment analysis has become the preferred method for categorizing text due to its accuracy and efficiency in determining whether a message is positive, negative, or neutral. It is crucial to identify which particular aspect or characteristic of a brand, product, or service a user is talking about to gain actionable insights.

Numerous instances of policy execution have resulted in public protests, including the Citizenship Amendment Act of 2019 (CAA) (Bhatia & Gajjala, 2020), the Agniveer Scheme of 2022, the Farm Bills of 2020 (Narayanan, 2021), and others. This shows the need to thoroughly understand public opinion before enacting any rule or policy. As a result, it is critical to create a framework that can be used by the government and non-governmental organisations to assess people's perspectives before enacting policies or laws for residents.

The influence of digitalization and virtualization is particularly noticeable in the realm of online education. More and more people are turning to the internet to receive formal education nowadays. The way children were taught underwent significant changes after the pandemic in comparison to the past. The outbreak of the novel Coronavirus disease prompted a shift towards virtual and online platforms, which allowed for interactive discussions and participation (Gaur et al., 2024) Kaur et al. conducted a case study on engineering students' attitudes toward online learning during the COVID-19 pandemic. Based on feedback from 123 students who participated in the online survey, it was found that online learning is likely to be used as a complementary approach alongside traditional classroom instruction, rather than as a standalone strategy (Kaur et al., 2021). When the virus was spreading and causing devastation, educational institutions were compelled to close. Countries were implementing lockdowns to prevent the virus from spreading. Online education and e-learning tools proved useful during that period. In addition, the National Education Policy 2020 (Jain, 2021) places great importance on online platforms like Massive Open Online Courses (MOOCs) and the

National Academic Credit Bank (Vashistha et al., 2022). As a result, a new standard has emerged where students have the option to attend classes and earn credits through online courses.

Massive Open Online Courses (MOOCs) are typically utilized for advanced education and professional growth. Coursera, Udemy, edX, Udacity, FutureLearn, and Khan Academy are some of the best online places to find high-quality MOOCs. These platforms collaborate with prominent universities and organisations to provide a diverse range of courses in computer science, business, engineering, humanities, and social sciences. Many of these platforms also offer professional certificates or degrees that learners can earn by completing a series of courses. Students enrolled in MOOCs can collaborate and create study groups by exchanging comments on various forums and asking and answering questions on social media sites and blogs. This information is crucial to the success of MOOCs, as it can be analysed to gain insight into student feedback and make modifications that enhance student participation and contentment. Although there may be a significant number of registered users, and even if the majority of them join merely for exploratory purposes, there may be a large number of forum posts or product reviews, and reading and evaluating them all individually would be impossible.

Numerous investigations have been conducted to assess MOOCs by scrutinizing the reviews associated with these courses. In their study, Jain and Gupta ((Jain & Gupta, 2022) explored MOOCs and mobile learning. They found that e-learning and mobile learning are emerging technologies. MOOCs offer video-based learning but lack instructor engagement. Mobile learning includes live streaming and social networking. Chakraborty's study explored student perspectives on online lectures, revealing positive feedback and interest in hybrid courses. Among 35 undergraduate computer science students, 43% found online lectures equally effective, 49% appreciated their flexibility, and 54% acknowledged improved teaching skills by professors. Additionally, 77% expressed interest in future online-offline hybrid courses (Chakraborty, 2022) .

Rdouan Faizi introduced a machine learning methodology to categorize the sentiment of learners' feedback on education-related videos on YouTube (Faizi, 2022). The feedback was classified as positive or negative using conventional machine learning approaches, including random forest, logistic regression, Naïve Bayes, and SVM. While each algorithm may have its own specific limitations, the SVM approach showcased exceptional results in this study. Ya Zhou (Zhou et al. 2020) investigated and compared the efficacy of online courses using BERT, XGBoost, and SVM models fused with binary characteristics in their study. It was demonstrated that when the BERT model was fused with binary characteristics, its performance improved when compared to using only the BERT model. Mingming Zhou (Zhou et al., 2022), analysed people's perspectives in China on online education during COVID-19 times using Excel software, the BERT pre-trained model, and the Latent Dirichlet Allocation algorithm.

MOOCs played a crucial role in reshaping the landscape of education by making it more accessible, affordable, and adaptable to the needs of learners worldwide. The paper addresses challenges in handling extensive data due to widespread digital technology adoption. The objectives of the study are as follows:

- Create a methodology for analysing reviews, identifying patterns, and gaining insights into user behaviour in order to extract valuable data.

- Develop a machine learning framework using sentiment analysis and k-means clustering. The framework seeks to identify and extract sentiments and trends represented in customer reviews using real-world case data.

- Describes an application of the proposed method to assess the quality of MOOCs.

The proposed framework has a wider application, as it may be used not only for product reviews but also for the assessment of educational resources like MOOCs. Data from Udemy and Coursera course reviews on MOOCs is pre-processed and annotated. Preprocessing eliminates unnecessary elements, and Name Entity Recognition (NER) improves accuracy. The BERT model is trained on this data set. This trained model identifies subsequent positive and negative ratings for the prediction dataset. Utilizing the K-mean clustering algorithm, the optimal number of positive and negative data clusters is determined. The quality of each cluster is then determined by computing the frequency of its top 10 words.

The document is organized as follows: The literature overview is offered in Section II, and the methodology and framework for a quantitative approach to the evaluation of user sentiments on real-world case studies of MOOCs are outlined in Section III. Section IV evaluates the proposed framework, and presents the results and findings, while Section V presents a comparative study of the proposed framework with other existing solutions. In Section VI, the future prospects of the research work are discussed.

2. Literature Survey

People now a days discuss their emotions and other personal experiences in an open and honest manner on a variety of online platforms. This has led to the creation of a huge and exquisite source of data that has caught the attention of researchers. Supervised Deep Learning (DL) and Machine Learning (ML) methods have been used in a number of studies to figure out what people think about controversial issues (Thukral et al., 2021; Varshney et al., 2021; Singh et al., 2021; Bilal & Almazroi, 2022,). Researchers have examined the efficacy of the BERT model as a means of representing data in texts. The BERT model can be explored further with the help of embedded clustering methods like eigenspace-based fuzzy c-means, deeply embedded clustering, k-means clustering, and deeply embedded clustering (Subakti et al., 2022). The study found that the BERT model yielded positive results with enhanced performance. Moodi and Saadatfar used a creative rendition of the K-means grouping technique, which increased clustering quality for large datasets (Moodi & Saadatfar, 2022). The proposed solution outperformed the A-means algorithm by 41.85 % in Shuttle and 41.15 % in quality when compared to the K-means methodology. Dahiya et al. used a weighted k-means clustering algorithm to study the online social behavior of WhatsApp users and categorize the chats' emotions (Dahiya et al., 2022).

Medford employed Latent Dirichlet Allocation (LDA) model for examining the tweets' data on coronavirus and determined the most common emotions of the people expressed. The tweets that were retweeted were used to decide the public common expression to determine emotion, and polarity (Medford et al., 2020). The findings indicated that roughly 30 percent of tweets expressed shock and 49.5 percent exhibited dread.

In recent research, Alhajji et al. employed the Naive Bayes classifier to examine and categorize the data scraped from Twitter and deduced that majority of tweets were positive regarding COVID-19 precautions (Alhajji et al., 2020).

Rajput et al. examined the emotions expressed in Twitter posts during the Coronavirus epidemic. Their work focused on word occurrence patterns and sentiment analysis utilizing the inbuilt TextBlob package (Rajput et al., 2020). The tweets were classified into positive, negative and neutral classes. The results showed that most tweets were positive, with very few expressing negative sentiments, which is indicative of the optimistic outlook of the population during the outbreak.

Kaila et al. used the LDA model to examine the information flow on Twitter during the COVID-19 outbreaks. Topic Modeling was applied to the document word matrix produced from the text corpus using LDA and the Gibbs sampling technique. Their research indicates that the novel coronavirus that is sweeping the world is causing anxiety in people, and that tweets regarding healthcare are frequently posted on social media (Kaila et al., 2020).

Kaur and Sharma examined public sentiment on coronavirus using TextBlob and the NLTK packages (Kaur and Sharma, 2020). Their research revealed that the majority of tweets contain the neutral opinions of individuals. This demonstrates additional room for improvement in the sentiment analysis of Twitter data.

Chen et al. examined user experience both before and after the COVID-19 outbreak using a comprehensive evaluation approach (Chen et al., 2020). Users' concerns about the online learning platform were analyzed in terms of course management, communication and engagement, learning, technical support, access speed, dependability, and timely delivery of video information. The study's findings showed that, prior to COVID-19, Zoom Cloud, Tencent Meeting, DingTalk, MOOC, TIM, WeChat Work, and Chaoxing Learning were ranked highest to lowest in terms of user experience. Following the COVID-19 pandemic, the following platforms were ranked in descending order of user experience: DingTalk, Zoom Cloud, Tencent Meeting, WeChat Work, MOOC, TIM, and Chaoxing Learning.

The efficacy of e-learning in vocational courses for students during the coronavirus outbreak was examined by Muktiarni et al. Students' comments and views through a questionnaire (Muktiarni et al., 2022) were collected. According to the results of the study, the majority of students intended to participate in online learning, but found it difficult to focus on academics as the teacher assists students to ensure that the learning process achieves the desired results.

Kaur investigated and evaluated the challenges and opportunities encountered during COVID-19, as well as the effects of the pandemic on the education sector (Kaur, 2021).

AppFollow software was used to perform a comparative analysis between the SWAYAM and Coursera in order to identify user sentiments, emotions, and feelings (Kaur et al., 2022) using the Sentiment Analysis. Reviews for both SWAYAM and Coursera were gathered between October 3, 2016, and October 3, 2021. A total of 40.5K ratings for SWAYAM and 127.7K ratings for Coursera were obtained.

Aytuğ ONAN utilized the sentiment categorization method on MOOC reviews using principles of machine learning, ensemble learning and deep learning (ONAN, 2021). Long-short-term-memory (LSTM) networks with GloVe word embedding exhibited the greatest classification accuracy of 95.80% in all configurations utilized in the work.

J. Shailaja and Sandhya assessed MOOC quality and its impact on students pursuing higher education using a questionnaire. They found that variability and interactivity are critical in MOOCs. To enhance student motivation and learning quality, MOOC providers should adopt innovative technology-based teaching methods (Shailaja & Prathikantham, 2018).

Ruba et al. used online lab demonstrations to assess students' satisfaction with COVID-19. Students appreciated in-person interactions with professors and equipment. The study suggests a hybrid laboratory approach with flexible hours and hands-on training to provide students with valuable skills for future employment (Alkhasawneh, 2024).

Vinay Kukreja et al. investigated the elements affecting online education. They collected responses from 673 individuals across various Indian colleges. Through exploratory and confirmatory factor analyses, they identified key factors influencing student satisfaction. . The findings demonstrated that instructor quality, course design, ICT orientation, conscientiousness, open-mindedness, and agreeableness all had a positive impact on satisfaction, whereas extraversion had a negative impact (Kukreja, 2021).

To assess participants' experiences in MOOCs, Xiaowei Yan et al. conducted a sentiment analysis employing the Naïve Bayes classifier (Yan et al., 2021).

In order to evaluate the experience of participants on the MOOC website of China University. Xiaowei Yan and colleagues conducted a sentiment analysis using the Naïve Bayes classifier (Yan et al., 2021). The findings indicate that while students report dissatisfaction with the assessment, they feel extremely satisfied with the course instructor, interactions, and content.

Although various studies have been reported on investigating users' responses to Online education, the quantification of the emotions and sentiments of people for assessing the quality of MOOCs is largely unknown. This work employs the BERT model and K-means clustering algorithm to quantitatively estimate the quality of the online courses before selection.

3. Research Methodology

To prevent the disastrous effects of divergent viewpoints, it has become crucial to assess the public's opinions on current issues. Modern times are characterized by rapid progress and the introduction of several new schemes and strategies across numerous industries and sectors. To determine their usage and importance in light of people's acceptance, it is necessary to gauge and assess public opinion on the newly introduced schemes and initiatives launched by the government or any formal authority. It enables the authorities to take actions that reflect public opinion and lead to less disagreement in society. Therefore, this work proposes a framework to address the aforementioned issues by analyzing and assessing public opinion.

Initially, public opinions from any social media platform or online repositories are gathered and categorised into positive and negative classes using the BERT-Bidirectional Encoder Representations from Transformer model, as shown in Fig. 1. The BERT model reads the entire string of words simultaneously, from left to right and right to left. This trait enables the model to learn a word's context based on its surroundings.

The BERT Model comprises of the architecture discussed as follows:

i. Input Layer (128 tokens): This represents the initial input layer with 128 tokens, indicating a sequence length of 128.

ii. Self-Attention and Feedforward Layers (Layers 1–12): Each layer in the BERT encoder consists of self-attention and feedforward sub-layers. The model is able to capture contextual information by weighing the relevance of individual words in the input sequence using the self-attention mechanism. To discover intricate relationships within the sequence, the feedforward layer analyses the results of self-attention.

iii. Layer Normalization and Residual Connections: Training is stabilised by the normalisation layer, which normalises the output of the sub-layers. Deep networks can be trained more effectively by using residual connections, which facilitate the gradient's flow during backpropagation.

iv. Pooling Layer: Information from the output of each encoder layer is combined by the pooling layer, typically using a pooling technique like mean or max pooling. As a result, the input sequence is converted into a fixed-size representation, also known as a pooled representation, which is used for activities later on.

v. Fully Connected Layers (Logits): The pooled representation is fed into the fully connected layers, which then turn it into the desired output. The output of BERT could be applied to a number of tasks, such as question-answering, named entity recognition, and text classification. Depending on the job, the output layer's activation varies; for example, the SoftMax is used for classification tasks.

Overall, this architecture uses many layers of feedforward processing and self-attention to extract contextual information from the input sequence. This results in a pooled representation that is subsequently utilised for downstream tasks through fully connected layers.

The BERT model is used to categorise the reviews gathered from the internet corpus, online repositories, or surveys, based on positive and negative sentiments expressed by the people. Subsequently, the K-means clustering algorithm is applied to find semantically similar clusters to determine and assess the quality of the system under study. Thereafter, the final intensity value reflecting the quality of the system is computed.

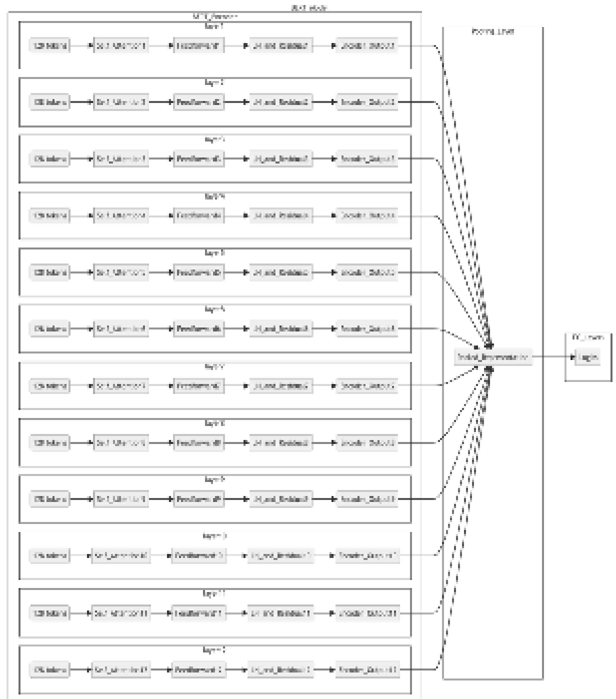


Fig.1: Architecture of the BERT model.

The flowchart of the proposed framework for evaluating and assessing the reviews is depicted in

A. Data Collection

The collection of acceptable corpora is critical for obtaining the desired and most relevant findings. It entails obtaining raw data or reviews from online repositories or social media platforms. The reviews are pre-processed to ensure that they are suitable for further processing.

B. Data Pre-Processing

Data gathering is followed by pre-processing. The reviews are pre-processed to remove unnecessary and undesirable data from the original data for effective model training and processing for accurate results. Data pre-processing entails removing special characters such as URLs, punctuation, hashtags, emoticons, emojis, HTML references, and numeric characters.

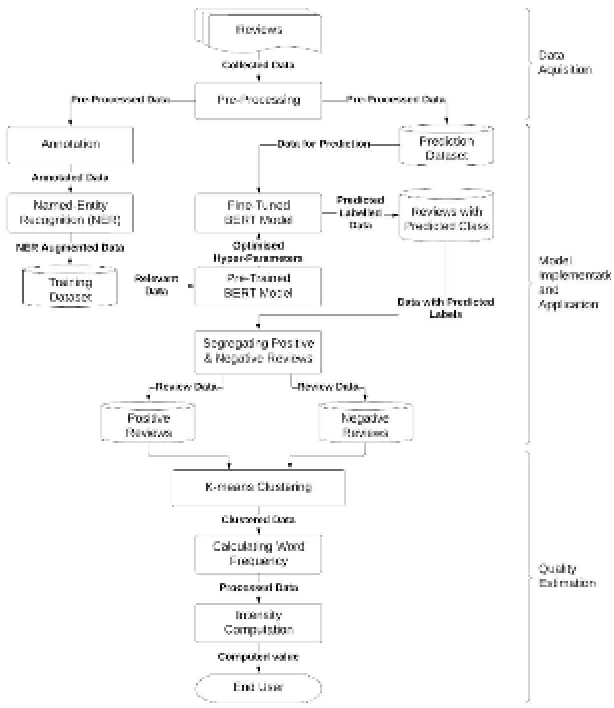


Fig. 2 : Proposed Framework for Assessing the Quality

C. Annotation

The process of labelling a dataset for categorization is known as data annotation. Furthermore, annotation of the collected reviews for the study is necessary to train the model based on the desired outcome. To acquire impartial and unbiased annotations, the participation of more than one annotator is desired. In this study, the reviews are classified and labelled into two groups: Positive (1) and Negative (0). The reviews that contained positive sentiments and good viewpoints are classified as positive, while the others with inappropriate opinions are marked as negative.

D. Named-Entity Recognition (NER)

Named Entity Recognition (NER) is a method that recognizes and classifies textual items, including places, names of specific people, organizations, and more. The precise identification and classification of these entities by NER contribute to an enhanced understanding of text and improve analytical capabilities. Consequently, various artificial intelligence (AI) applications, such as chatbots and sentiment analysis, benefit from NER.

Although NER has several advantages, its implementation is challenging. Managing entities that are uncertain or have numerous plausible categorizations is a serious difficulty. Precisely differentiating between these entities requires intricate methods and a thorough analysis of the surrounding context. Furthermore, NER has difficulty distinguishing named entities in languages other than English due to potential constraints in training data. Language variances, misspellings, and the inclusion of previously undiscovered or novel named entities not present in the training data can have an adverse effect on NER performance. Advanced deep learning models, such as transformers and the BERT model are utilized in this work to address these challenges. The models go through significant pre-training on a large scale before being fine-tuned using subject-specific data to produce the best outcomes. Further advancements in NER are possible as a result of the utilization of transfer learning and the development of pre-trained models tailored to specific domains.

E. Training and Testing of the Model

The study utilizes the BERT model, distinguished by its architecture as a stack of transformer encoders. BERT employs a multi-layer bidirectional transformer encoder to efficiently handle substantial amounts of textual input and accurately capture intricate word relationships. Bidirectionally trained language models have a better comprehension of language context and flow than single-direction models, resulting in better results.

In conjunction with its transformer encoders, BERT employs a two-step approach comprising pre-training and fine-tuning. During the pre-training phase, BERT is trained to predict missing words in sentences using an extensive dataset of unlabeled text. Through this pre-training phase, BERT attains a comprehensive understanding of language and context. Subsequently, in the fine-tuning stage, BERT is trained on labelled data for specific tasks, such as question answering or sentiment analysis. This fine-tuning process enables BERT to adapt to individual tasks, enhancing its performance across a broad spectrum of Natural Language Processing (NLP) applications.

The BERT-base-uncased model is optimized for binary classification (Devlin et al., 2018). The model was trained using the pre-processed annotated reviews. The encoder read the tokenized texts bidirectionally to determine the context of the token.

This bidirectional training enabled the model to comprehend the language context completely and without ambiguity.

F. Segregating Positive and Negative Reviews

After applying the fine-tuned BERT Model to the dataset, reviews with predicted labels are obtained. For further evaluation, these reviews are divided into positive and negative categories. The segregation is done to assess the quality of reviews.

G. K-Mean Clustering

K-Means Clustering is an unsupervised learning method that divides unlabeled data into clusters. The number K represents the number of pre-defined clusters to be produced. Algorithm 1 below describes K-means clustering in brief.

The K-means clustering technique separates the dataset into clusters of semantically comparable phrases. The K-mean technique is applied to obtain the positive and negative reviews separately. The text in the clusters is used to draw word clouds to find out the most significant words in the clusters.

H. Word Frequency

The Word frequency of each cluster is determined to explore the most frequently appearing words in the clusters obtained in the preceding step. It is done using Python data analysis and visualization tools, like Matplotlib, to find the top 10 most frequently occurring words in the clusters.

I. Computing Quality

Algorithm 2 mentioned below is used to find Intensity value of the top 10 words from each cluster for computing the Mean Quality Value (MQV) of the system.

Algorithm 1: K-Means Clustering Algorithm

Input: Number of desired clusters K, Data points N
Output: A set of K clusters
Step 1: Randomly initialize k centroids.
Step 2: Each data point in N should be connected to the closest centroid. The data points will be split into k clusters as a result.
Step 3: Recalculate the centroid's positions. repeat steps 2 and 3 until the membership of the data points has not changed any more.

Algorithm 2: To find the Quality of the cluster

Let S be an array of top 10 words w_i *where i ranging from 0-9 in each cluster.*
Let F be the frequency of each word in S.
Let W be the weight of w_i *in a cluster.*
Let $T = \sum_{i=0}^{9} F[w_i]$ *be the total sum of the frequency of the top 10 words.*
Let Q be the quality value of w_i *in cluster*
For each w_i *in S:*
    $W[w_i] = F[w_i]/T$
For each review in the cluster:
    Let Count = 0
        For every word in the review of the cluster:
      If a word is in S:
        Count = Count ∪ weight of W[word]
    Q = Q ∪ Count
Q = Q / Number of reviews in a cluster

Algorithm 2 begins by defining an array S comprising the top ten words (w_i) in each cluster, as well as their frequency (F) and weight (W). The weight of each word (w_i) in a cluster is calculated as the ratio of the frequency of the word (F[w_i]) to the total sum of the frequency of the top 10 words (T). This provides a normalized representation of the importance of each word in a cluster.

The algorithm then iterates over each review in the cluster. For each review, the algorithm calculates the weight of all the terms in the review that appear to be in the top ten (S). The count for each review is added to calculate the cluster's quality value (Q). Finally, the quality value (Q) for the cluster is derived by dividing the total count by the number of reviews in the cluster. This provides a measure of the overall quality of the cluster based on the weight of the terms in the reviews.

Results and Findings

The outbreak of COVID-19 forced the closure of all educational facilities, putting a spotlight on the effectiveness of online and digital education. When everything seemed to be coming to a halt, online education technologies proved to be the only way to impart knowledge and engage students and educators (Devlin et al., 2018). Therefore, it becomes essential to look into online education and find out what stakeholders think about it.

The proposed framework is utilized to evaluate the broad perspective of public opinion regarding online education, as well as their experiences with it. A dataset from www.kaggle.com was used to implement the proposed framework. It is an online community for data scientists and machine learning engineers that provides a platform for discovering datasets for Artificial Intelligence (AI) model development, publishing datasets, collaborating with peers, and participating in data science competitions. Kaggle upholds a commitment to privacy by not selling personal data, as stated in the "Sharing Your Information" section of their privacy policy. Consequently, this study was carried out with ethical integrity, with a focus on the privacy of reviewers.

The dataset included reviews on Online Education from MOOCs such as Udemy and Coursera with ratings ranging from 5 stars to 1 star, with five stars being the highest and one star being the lowest. A total of 3,000 reviews were used for the study, of which 2,500 were used for BERT model training and 500 were used for testing.

The dataset was processed to remove duplicates and non-English notices. Before training the BERT model, the dataset was processed to remove any special characters, punctuation, hashtags, and URLs. Furthermore, the dataset was processed to remove stop words that were not useful in training the BERT model. As the names of the people, and the date of the reviews, and the course of study were mentioned in the reviews, it was essential to replace them with their corresponding category types to avoid information bias during model training. So, the dataset was additionally processed using NER, which involved identifying crucial information in the text and classifying it into a set of predefined categories such as CARDINAL, DATE, TIME, PERSON, ORG, and GPE to improve the efficiency of the BERT Model. These categories represent the numeral, absolute or relative date, time shorter than a day, people, including fictional names, organizations, and states, respectively. In addition, this study included three independent subject experts who annotated the collected reviews. The guidelines for annotation involved the following recommendations; reviews with a best-to-good rating received a rating of 1(positive), reviews with a worst-to-bad rating received a rating of 0 (negative), and everything else was dropped.

The BERT base-uncased model was used to carry out the classification of the pre-processed reviews. Python's Transformers and TensorFlow 2.0 libraries were used to implement the BERT model. The pre-trained BERT model was adjusted for the binary classification of reviews. To avoid overfitting, a dropout layer with a 50% dropout rate was added to the BERT model. Thereafter, a Dense layer of 768 neurons was added with the Tanh activation function, which was followed by a second dropout layer with a 50% dropout rate, and an additional Dense layer of two neurons was used with the SoftMax activation function. Based on the probability scores of the BERT model, the reviews were divided into two broad categories: positive and negative. The BERT model utilized a performance optimization loss function, the Sparse Categorical Cross Entropy. The training phase was repeated with various values of hyperparameters like the optimizer function, epochs, number of layers, and learning rate. The model was eventually adjusted with the final values of hyperparameters as shown in Table 1.

Table 1 : Final Values of Bert Hyper Parameters

Hyper-Parameters	Optimal Value
Optimizer	Adam
No. of Layers	12
Max. Sequence Length	128
Learning Rate	3e-5
No. of Epochs	2
Loss Function	Sparse categorical cross entropy
Batch Size	16
Dropout	0.5

The model was used to make predictions after it had been trained and tuned. The metrics listed below are used to evaluate the model's performance (DalianisHercules D, 2018):

- Accuracy refers to the proportion of accurately predicted observations as compared to the total number of observations.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- Precision refers to the proportion of accurately predicted positive observations among the total projected positive observations.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- Recall refers to the proportion of accurately predicted positive observations out of all observations in the actual class.

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score is a metric that computes the weighted average or the harmonic mean of Recall and Precision.

$$F1 - score \ = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Here

TP - Ture Positives
FP - False Positives
TN - True Negatives
FN - False Negatives

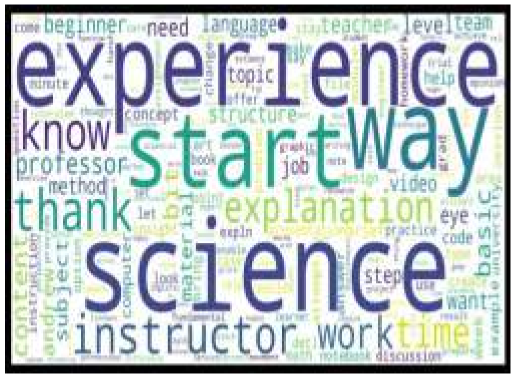
The higher the values of the aforementioned metrics, the more reliable the model is at making predictions. The reliability of the model improves as the value of the metrics increases. The model obtained an accuracy of 95.5% on the annotated reviews test set, with precision and recall rates of 94.5% and 95.4%, respectively, and an F1-score of 95%.

After training, the BERT model was used to make predictions. Reviews with a probability score of over 0.5 were considered positive, and reviews below 0.5 were considered negative. Both positive and negative reviews were processed separately through the K-means algorithm to find semantically similar information in the cluster. The optimal number of clusters was determined by the 'elbow method'. It is a technique for determining the ideal number of clusters for K-means clustering. 'K' represents the optimal number of clusters.

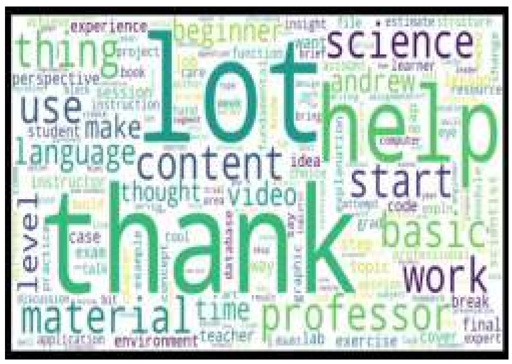
Three optimal clusters were obtained for the positive reviews and five for the negative reviews. Further, clustered reviews were used to assess the quality of the online courses.

The clusters are represented using word clouds. Figs 3(a), 3(b) & 3(c) show the word clouds of the positive clusters, while and Figs 4(a), 4(b), 4(c), 4(d) and 4(e) display the word clouds of the negative clusters. The size of the word in a cloud indicates the word's closeness to the cluster's centroid. The clusters of positive reviews indicated the good and desirable characteristics of MOOCs, while clusters of negative reviews may reveal the context of the improvements.

After categorizing the reviews into positive and negative clusters, each cluster was analyzed to determine the valuable information offered by the reviews. This was accomplished by identifying the top ten words in each cluster and their frequencies. Figs 5(a), 5(b), and 5(c) depict the top ten words of the positive clusters, while the top 10 words in the negative clusters are shown in Figs 6(a), 6(b), 6(c), 6(d) and 6(e). To quantify quality, the top 10 words were picked up from each cluster based on their frequency, as shown in Table 2. And the weight of these top words was calculated by dividing the frequency of the word by the total occurrence of all the top 10 words. Thereafter, reviews in each cluster were analyzed to determine the overall quality of the cluster. Each review was scanned from left to right to find out if it contained any of the top 10 words.



(a) Cluster 1



(b) Cluster 2

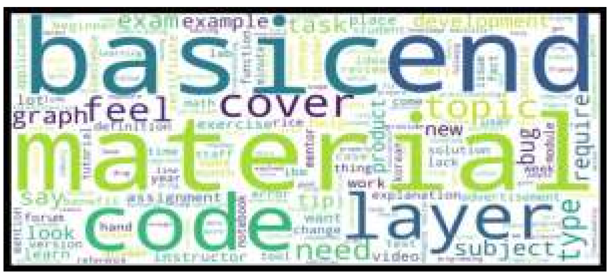
(c) Cluster 3

Fig. 3 : Word Cloud of Optimal Clusters for Positive Labelled Reviews

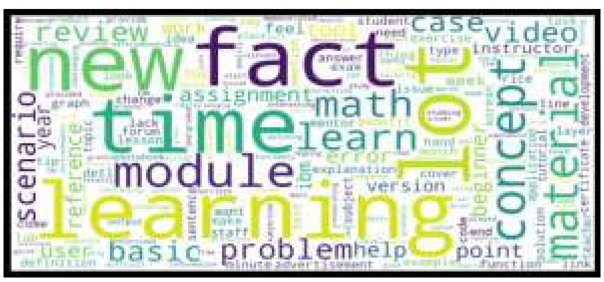
(a) Cluster 1

(b) Cluster 2

(c) Cluster 3



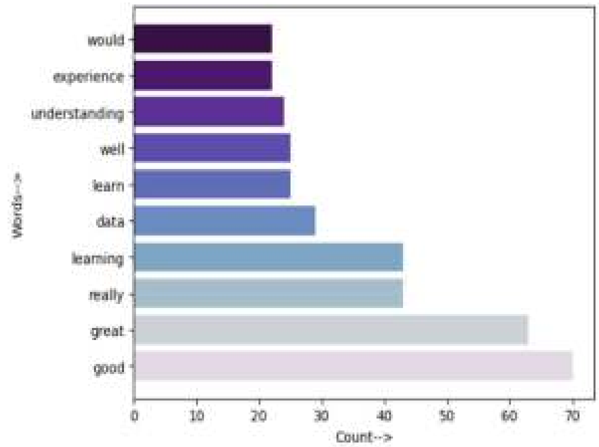
(d) Cluster 4



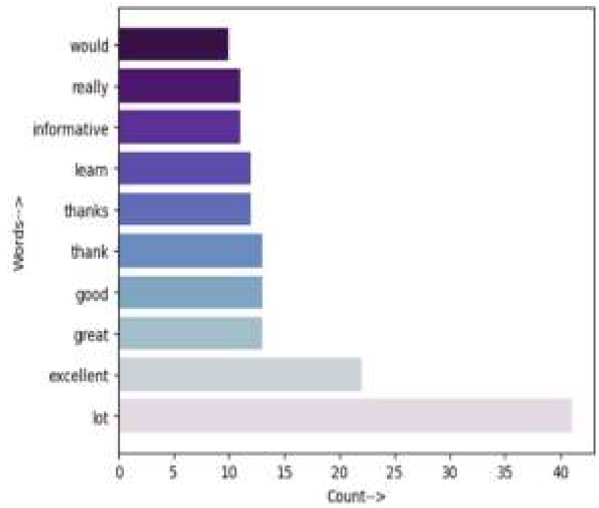
(e) Cluster 5

Fig. 4 : Word Clouds of Clusters for negative labelled reviews

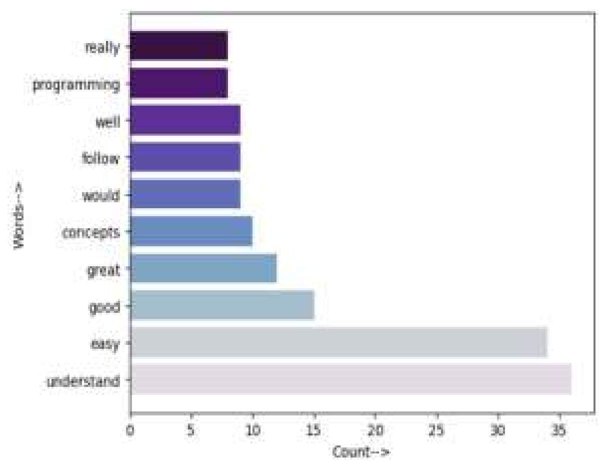
Positive Clusters:



(a)Top 10 words for positive cluster 1



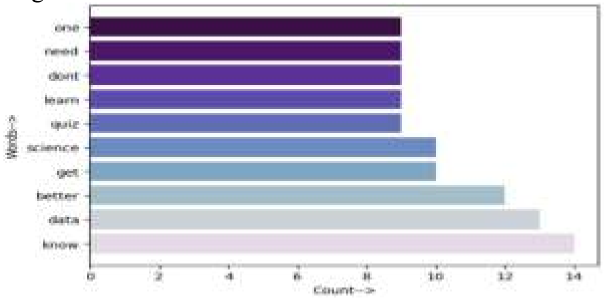
(b)Top 10 words for positive cluster 2



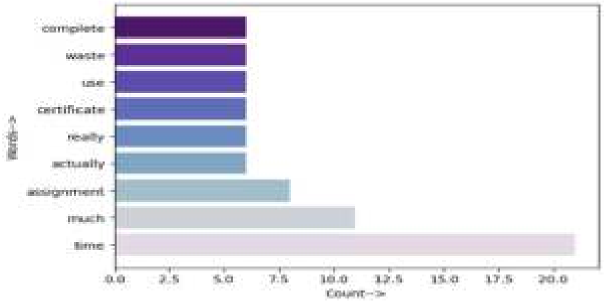
(c)Top 10 words for positive cluster 3

Fig. 5 : Graphical Representation for Top-10 words in Positive Clusters

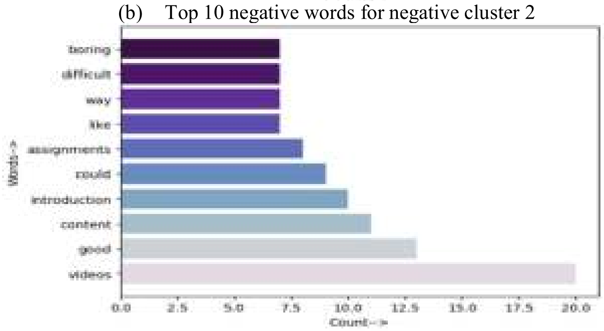
Negative clusters:



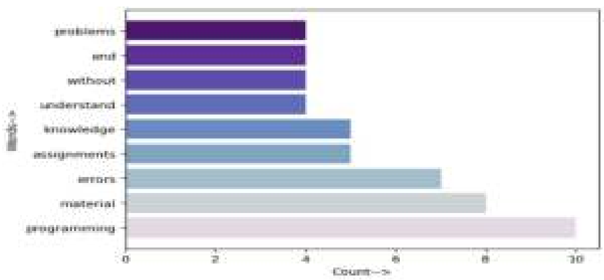
(a) Top 10 negative words of negative cluster 1



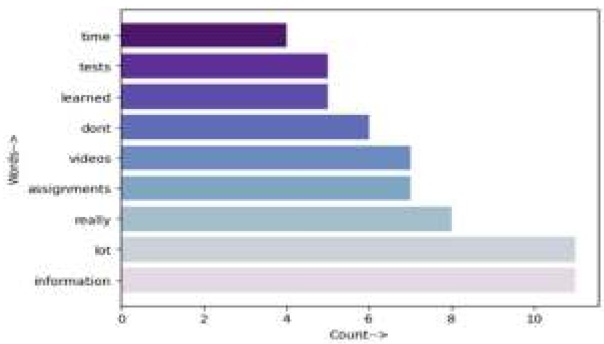
(b) Top 10 negative words for negative cluster 2



(c) Top 10 negative words for negative cluster 3



(d) Top 10 negative words for negative cluster 4



(e) Top 10 negative words for negative cluster

Fig. 6 : Graphical Representation for Top-10 Words in Negative Clusters

Table 2 : Frequency of Top 10 Words in the Cluster

Word	Frequency
'good'	70
'great'	63
'really'	43
'learning'	43
'data'	29
'learn'	25
'well'	25
'understanding'	24
'experience'	22
'would'	22

Table 3 : Mean Quality Value of Each Cluster

Cluster	Mean Quality Value	Quality Characteristics Exhibited by a Cluster
Positive Cluster 1	0.172	Excellent and easy to understand
Positive Cluster 2	0.252	Content is good and enjoyable
Positive Cluster 3	0.360	Teachers were experienced, excellent introduction and impressive teaching method
Negative Cluster 1	0.167	Not enough examples
Negative Cluster 2	0.316	Too many assignments and lacks detailed explanation
Negative Cluster 3	0.233	Missing exercise questions
Negative Cluster 4	0.229	Too long and doesn't offer real world knowledge
Negative Cluster 5	0.247	Very time consuming

The frequency of the word was then multiplied by its weight to compute the quality contribution of the review in a cluster. The process was repeated for all the reviews in a cluster, and the quality values of each review were added to obtain the final quality value of the cluster. The quality value of a cluster was divided by the cluster's number of reviews to determine its Mean Quality Value (MQV). Table 3 displays MQVs for each cluster.

The MQV represents the degree to which reviews within a cluster exhibit a particular quality attribute. A lower MQV indicates that the course has fewer quality characteristics.

The MQV for Positive Cluster 1 was 0.172, or 17.2 percent, indicating that the course is excellent and simple to understand. As evidenced by the MQV of 0.252 for Cluster 2, the audience liked and appreciated the course material. Positive Cluster 3, with a MQV of 0.36, indicated that the instructors were experienced and that the course gave an excellent introduction to the subject. Overall, reviews of positive clusters led to the conclusion that MOOCs present the content clearly and succinctly while covering the fundamentals. These courses have experienced instructors with impressive teaching strategies. Additionally, MOOCs offer the flexibility to select the course.

Conversely, Negative Cluster 1, shown in Figs 4(a) and 6(a), with a MQV of 0.167, demonstrated that MOOCs fall short in terms of providing enough examples to teach the subject. Figs 4(b) and 6(b) of Negative Cluster 2 with MQV 0.316 demonstrated that the course content was introductory and that there were a few assignments in the course. As shown in Figs 4(c) and 6(c), Negative Cluster 3 with MQV 0.233 indicated that learners wanted more questions following each class to gain a better comprehension of the subject. Negative Cluster 4, with a MQV of 0.228, revealed that the course was too lengthy and lacked essential real-world knowledge, as depicted in Figs 4(d) and 6(d). Finally, Negative Cluster 5, with an MQV value of 0.247, indicated that the course was time-consuming and the content offered was not advanced, as can be observed from Figs 4(e) and 6(e).

It may be concluded from the study that MOOCs are effective for basic subjects with knowledgeable instructors. However, they lack sufficient examples, advanced knowledge, and real-world experience. Students desire more questions after lessons, and the courses are lengthy.

The Analytical Hierarchical Process (AHP) was employed to assess and validate the user's preferences for an effective education system, which could be online or offline. AHP is among the most prevalent techniques used for multi-criteria decision-making (MCDM). It is beneficial in identifying a viable alternative for a specific complex situation (~~Enrique~~ Mu et al.., 2017).

Features of the Online mode were extracted from the proposed research work that are listed below:

(1) easy to understand,

(2) Comfortable learning

(3) flexibility to choose a course or faculty

A survey was held with 21 academicians and 32 students to identify the following features of offline mode:

(1) Face-to-face interaction

(2) Candid environment of learning

(3) Rigid time constraints

A case study was conducted to compare the performance of online mode and offline mode using AHP (Mu, E. et al., 2017) with the following requirements of students:

(1) Basic Course

(2) Less time-consuming

(3) Interactive

AHP was utilized to evaluate the online or offline education as per the requirements. The results indicated that online mode could fulfill the desired specifications.

5. Comparative Analysis

These days, people often share their views and experiences on social platforms like Facebook and Twitter. They can openly and comfortably share their opinions and beliefs. Several machine learning and deep learning techniques have been used to measure public opinion on critical issues. Methodologies such as sentiment analysis, LSTM, BERT, K-means, and word frequency are employed in several studies.

Li et al. performed sentiment analysis on MOOC comments using the shallow BERT-CNN model as the comment classifier. The model consisted of a shallow pre-trained BERT (6 layers), a convolutional layer, and a self-attention pooling module (Li et al., 2019). Their model achieved an accuracy of 81.3% and an F1 score of 92.8%. Additionally, the application of convolutional neural networks and the self-attention mechanism yielded comparable performance while reducing BERT parameters by half.

Zhou and Li studied and compared the efficacy of BERT, XGBoost, and SVM models paired with binary characteristics for evaluating the success of online courses. Recall, precision, and F1-score for BERT and BERT combined with binary representation of the text data were 95.88 percent, 95.92 percent, and 95.90 percent, respectively, and 96.75 percent, 96.78 percent, and 96.75 percent, respectively (Zhou et al., 2020). Text data was represented using binary values, where each characteristic or element is either present (1) or absent (0) and is merged or integrated in a way that allows the text to be classified into multiple categories or classes. The study indicated that incorporating binary characteristics into the BERT model resulted in improved performance, outperforming the BERT model alone.

Yan et al. conducted a sentiment analysis employing the Naïve Bayes classifier on reviews from the China University MOOC website to assess participants' experiences (Yan et al., 2021). The classifier achieved an average accuracy of 85.62%. As a result, the findings indicate that learners experienced high satisfaction with the course instructor, interaction, and course content. However, they expressed dissatisfaction with aspects such as course assessment and the health of the learning platform. This study underscored the applicability of sentiment analysis (SA) in educational research and its potential integration into online learning systems, enabling real-time analysis of student feedback.

These studies used methodologies such as Deep Learning and Machine Learning models, but they were not utilized comprehensively and collectively to generate significant results. The outcomes of the aforementioned studies, did not shed light on the quality of MOOCs.

The proposed framework utilizes the annotations from Coursera and Udemy reviews for evaluating MOOC quality. The NER-enhanced data was utilized for training and testing the BERT model, leading to improved performance and increased model reliability. The K-mean clustering algorithm was used to obtain positive and negative clusters of reviews. The positive cluster indicated the liking of the course by the users, while the negative cluster showed the undesirable aspects of the MOOCs. Furthermore, by combining NER and BERT, the proposed framework outperformed all other machine learning techniques in terms of accuracy (95.5 percent) and F1 scores (95 percent). Additionally, the proposed work also quantified the positive and negative characteristics of these courses to help novice users select them. The proposed framework may be used to provide feedback to the designers of online courses.

6. Conclusion

The proliferation of digital technologies across numerous industries has resulted in an explosion of data that has proven challenging to manage. In a broad range of the public and private sectors, the ability to analyze and mine data from massive datasets is gaining importance.

The COVID-19 pandemic had a significant impact on the schooling system, resulting in disruptions to the conventional method of teaching and affecting both students and educators. As a result, there has been a surge in popularity for online teaching and MOOCs. Therefore, it is important to gather people's opinions and viewpoints to better understand the impact and implications of the shifts in the education system. In this research, a quantitative framework is proposed to analyze online education. The study utilized sentiment analysis to evaluate user opinions and feedback on MOOCs. The reviews were categorized into positive and negative clusters to determine the overall positive and negative quality attributes of MOOC courses. The proposed framework can help individuals make informed choices before selecting the course. The results of the study indicated that MOOCs provide clear, concise content with experienced instructors using impressive teaching strategies, and also offer flexibility in course selection. Conversely, they don't provide sufficient examples, and questions following the lessons. These courses are lengthy and lack essential, real-world knowledge.

The work employed a comprehensive methodology to compute the quality of MOOCs, which expands the prospect of study for future implementation. The future possibilities of this research on MOOCs and online education span over various dimensions:

- Improvements in sentiment analysis methodologies may be realized through the integration of ML and AI advancements. This integration can be poised to refine the analysis of user feedback, offering more insights into both positive and negative sentiments.

- Collaborating with educators and subject experts to address the study's findings in MOOCs. This collaboration may provide scope for the refinement of course materials and the introduction of advanced-level content, ultimately improving the overall quality and depth of MOOCs.

- Conducting comparative studies among different

MOOC platforms or courses that can provide an opportunity to benchmark quality standards and identify best practices. These comparative analyses will empower learners to make more informed choices.

Exploring these avenues enables researchers and educators to elevate the quality, accessibility, and effectiveness of MOOCs, contributing to the continuous evolution and improvement of online education.

The proposed work can be extended and applied to various sectors and fields, such as the industrial sector, healthcare, and so on, to gain a better understanding of its working environment. Furthermore, the self-employed and business people could use this framework to undermine the people's outlook and achieve high standards by improving on the insights provided by the people. The study's findings will also be useful to the government in terms of the effective implementation of laws and policies designed for people.

Appendix - I

S. No.	Abbreviation	Description
1.	MOOC	Massive Open Online Course
2.	BERT	Bidirectional Encoder Representations from Transformers
3.	CAA	Citizenship Amendment Act
4.	NEP	New Education Policy
5.	DL	Deep Learning
6.	ML	Machine Learning
7.	NER	Name Entity Recognition
8.	CNN	Convolutional Neural Networks
9.	LSTM	Long Short-Term Memory
10.	Bi-LSTM	Bidirectional Long Short-Term Memory
11.	NB	Naive Bayes
12.	SVM	Support Vector Machines
13.	RF	Random Forest
14.	AI	Artificial Intelligence
15.	MQV	Mean Quality Value

Acknowledgment

The authors greatly acknowledge the facilities provided under the STAR College scheme, Department of Biotechnology, in carrying out this work. The authors are also grateful to the Departments of Computer Science and Electronics at Acharya Narendra Dev College, University of Delhi, for providing the necessary conceptual understanding and financial assistance to carry out this work.

References

Alhajji, M., Al Khalifah, A., Aljubran, M., & Alkhalifah, M. (2020). Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19.

Alkhasawneh, R., Al-Nsour, R., & Akho-Zahieh, M. (2024). Measuring Online Engineering Laboratory Satisfaction During COVID-19 Pandemic. Journal of Engineering Education Transformations, 37(3), 157–164.

Bhatia, K. V., & Gajjala, R. (2020). Examining anti-CAA protests at Shaheen Bagh: Muslim women and politics of the Hindu India. International Journal of Communication, 14, 18.

Bilal, M., & Almazroi, A. A. (2022). Effectiveness of Fine-Tuned BERT Model in Classification of Helpful and Unhelpful Online Customer Reviews. Electronic Commerce Research, 1-21.

Chakraborty, P. (2022). Opinion of Engineering Students on the Use of Live Online Lectures after the Pandemic. Journal of Engineering Education Transformations, 36(special issue 2), 148–155.

Chen, T., Peng, L., Jing, B., Wu, C., Yang, J., & Cong, G. (2020). The impact of the COVID-19 pandemic on user experience with online education platforms in China. Sustainability, 12(18), 7329.

Dahiya, S., Saluja, A., Singhal, P., & Johari, R. (2022). Analysis of Online Social Behavior of Whatsapp Users. Webology, 19(1), 229-249.

Dalianis, H., & Dalianis, H. (2018). Evaluation metrics and evaluation. Clinical Text Mining:

secondary use of electronic patient records, 45-53.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Faizi, R. (2022). A Sentiment-Based Approach to Predict Learners' Perceptions Towards YouTube Educational Videos. In: Abraham, A., et al. Innovations in Bio-Inspired Computing and Applications. IBICA 2021. Lecture Notes in Networks and Systems, vol 419. Springer.

Gaur, V., Bhangalia, A., Singhal, S & Kaur, R. (2024). A Smart Learning Management System for Virtual Labs. Journal of Engineering Education Transformations. 38(1), 164-177.

Jain, M. K., & Gupta, S. (2022). A Pragmatic Analysis of Mobile and MOOCs Based Learning Methods. Journal of Engineering Education Transformations, 35(4), 7–22.

Jain, P. K. (2021). A Critical Analysis of New Education Policy 2020 And It's Future Implications. Journal of Commerce and Trade, 16(2), 1-8.

Kaila, R. P., & Prasad, D. A. (2020). Informational flow on Twitter–Corona virus outbreak–topic modelling approach. International Journal of Advanced Research in Engineering and Technology (IJARET), 11(3).

Kaur, C., & Sharma, A. (2020). Twitter sentiment analysis on coronavirus using TextBlob (No. 2974).

Kaur, R., Garg, A., & Kaur, P. (2021). Case study: Student's response towards online learning in engineering education during COVID-19 pandemic. Journal of Engineering Education Transformations, 34(3), 62–69.

Kaur, S. (2021). Challenges And Opportunities in Education During Covid-19, GHG Journal of Sixth Thought Vol. 8, No. 1&2 March & Sep. 2021

Kaur, A., Sharma, R., Gulati, S., & Chakravarty, R. (2022). Mobile app analytics for assessing MOOC platforms: a study of online learners' sentiments. Library Hi Tech News.

Kukreja, V., Kaur, A., & Aggarwal, A. (2021). What factors impact online education? A factor analysis approach. Journal of Engineering Education Transformations, 34(Special Issue), 365–374.

Li, X., H. Zhang, Y. Ouyang, X. Zhang and W. Rong, "A Shallow BERT-CNN Model for Sentiment Analysis on MOOCs Comments," 2019 IEEE International Conference on Engineering, Technology and Education (TALE), Yogyakarta, Indonesia, 2019, pp. 1-6.

Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. (2020). An "Infodemic": Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak. Open Forum Infectious Diseases, 7(7).

Moodi, F., & Saadatfar, H. (2022). An improved K-means algorithm for big data. IET Software, 16(1), 48-59.

Mu, E., Pereyra-Rojas, M., Mu, E., & Pereyra-Rojas, M. (2017). Understanding the analytic hierarchy process. Practical decision making: an introduction to the analytic hierarchy process (AHP) using super decisions V2, 7-22.

Muktiarni, M., Ana, A., Barliana, M. S., Saleh, I., Mulyadi, Y., Budhi, A. & Muda, W. H. N. B. W. (2022). Readiness for Online Learning During Covid-19 Pandemic in Vocational Education. Journal of Engineering Education Transformations, 35(Special Issue 2).

Narayanan, S. (2021). Understanding farmer protests in India. Academics Stand Against Poverty, 1(1), 133-140.

Onan, A. (2021). Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. Computer Applications in Engineering Education, 29(3), 572-589

Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020).

Word frequency and sentiment analysis of twitter messages during coronavirus pandemic. arXiv preprint arXiv:2004.03925.

Shailaja, J., & Prathikantham, S. (2018). Survey on understanding the implications of MOOCs in engineering education. Journal of Engineering Education Transformations, 2018(Special Issue).

Singh, M., Jakhar, A. K., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. Social Network Analysis and Mining, 11(1), 1-11.

Subakti, A., Murfi, H., & Hariadi, N. (2022). The performance of BERT as data representation of text clustering. Journal of Big Data, 9(1), 1-21.

Thukral, T., Varshney, A., & Gaur, V. (2021). Intensity quantification of public opinion and emotion analysis on climate change. International Journal of Advanced Technology and Engineering Exploration, 8(83), 1351.

Varshney, A., Kapoor, Y., Chawla, V., & Gaur, V. (2021). A novel framework for assessing the criticality of retrieved information. International Journal of Computing and Digital Systems.

Vashistha, N., Dhiman, A., Grima, S., & Kiran, S. (2022) National Academic Credit Bank: Need of Times. In Edutech Enabled Teaching (pp. 201-216). Chapman and Hall/CRC.

Yan, X., Li, G., Li, Q., Chen, J., Chen, W., & Xia, F. (2021, October). Sentiment analysis on massive open online course evaluation. In 2021 IEEE International Conference on Neuromorphic Computing (ICNC) (pp. 245-249).

Zhou, Y., & Li, M. (2020). Online Course Quality Evaluation Based on BERT. 2020 International Conference on Communications, Information System and Computer Engineering (CISCE). https://doi.org/10.1109/cisce50729.2020.00057

Zhou, M., & Mou, H. (2022). Tracking public opinion about online education over COVID-19 in China. Educational Technology Research and Development, 70(3), 1083–1104. https://doi.org/10.1007/s11423-022-10080-5