THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 1 (2024)
https://doi.org/ 10.62019/abbdm.v4i1.132


# Machine Learning-Based Students' Sentiment towards E-Learning amid COVID-19 Pandemic

Fida Muhammad Khan*, Zahid Iqbal, Muhammad Shoaib Akhtar, Inam Ullah Khan

## Chronicle

**Article history**
**Received**: March 15, 2024
**Received in the revised format:** March 20, 2024
**Accepted:** March 23, 2024
**Available online:** March 27, 2024

**Fida Muhammad Khan and Zahid Iqbal** are currently affiliated with Department of Computer Science, University of Science & Technology, Bannu, Pakistan.
**Email:** fida5073@gmail.com
**Email:** s.zahidiqbal77@gmail.com

**Muhammad Shoaib Akhtar** is currently affiliated with Department of Electrical Engineering, University of Science & Technology, Bannu, Pakistan.
**Email:** engrshoaibkhanmrt@gmail.com

**Inam Ullah Khan** is currently affiliated with Department of Computer Science, University of Lakki Marwat, Pakistan.
**Email:** Inam1software@gmail.com

***Corresponding Author:**

## Abstract

The global COVID-19 pandemic has created exceptional difficulties for college and university students. With educational institutions shutting down and an abrupt shift to online learning, students are facing new challenges. Despite the advantages of e-learning, developing nations like Pakistan are experiencing major hurdles due to a lack of internet access and financial constraints. As a result, students' academic performance is being negatively impacted. Researchers are using sentiment analysis of social media platforms such as Facebook, Twitter, and YouTube to gain insight into student experiences during this challenging time. This study seeks to create a machine learning-based sentiment classification system that can determine the sentiment of higher education students in Pakistan towards e-learning classes during the COVID-19 pandemic. The primary objective of this research is to identify and categorize the challenges and problems that university students face in their e-learning classes. The research employed Multi-Nominal NB and Gaussian NB classifiers to examine the sentiment of online reviews and comments of students. We achieve 99% accuracy using Gaussian NB and 98% accuracy using Multinomial NB classifiers. According to the study's results, the COVID-19 epidemic has had a significant effect on the education system, with the vast majority of feedback being negative (70%). Only a small percentage of feedback was classified as neutral (29.9%) or positive (4%). This finding underscores the urgent need for better solutions for online learning during and after the pandemic. The study underscores the importance of addressing these unique challenges faced by students in developing countries during the pandemic. There is a pressing need for innovative solutions that can ensure that higher education remains effective and accessible in these challenging times. By utilizing sentiment analysis and machine learning, researchers can gain a better understanding of the experiences of students during this crisis. This understanding can inform targeted interventions and solutions that can improve academic performance and well-being for students in developing countries like Pakistan. The findings of this research provide important insights and recommendations for policymakers, educators, and stakeholders in the higher education sector.

**Keywords:** COVID-19, E-Learning Classes, Higher Education Commission of Pakistan, social media, Machine Learning, Classification, Gaussian Naïve Bayes, Multinomial Naïve Bayes.



# INTRODUCTION

Natural Language Processing (NLP) methods were used to do sentiment analysis on the data that was gathered through an online poll. The findings indicate that although students' attitudes about e-learning during the epidemic were generally negative, there were some reservations about internet connectivity, a lack of interaction with teachers and peers, and difficulties with time management

## Sentiment Analysis

The abundance of unrefined data available on social media can be a valuable resource for businesses. Thanks to technological advancements, particularly in the fields of machine learning and artificial intelligence, this data can now be analyzed and transformed into useful insights that have the potential to benefit a wide range of businesses (Abinash Tripathy et al.,2016). In today's world, social media platforms are widely used by people across the globe to share information. The users share their daily experiences and give their viewpoints on a variety of subjects including products and places. Companies can leverage this large-scale platform to collect information about user preferences. Some examples of social media platforms that offer global communication opportunities include Twitter, Facebook, and Instagram (Adnan et al.,2020). The increasing digitization of our world has led to a substantial increase in user-generated content on the internet.

This content includes people's opinions on various topics, which can be analyzed through a computational technique called sentiment analysis. By analyzing an entities emotions and attitudes towards a particular subject, sentiment analysis can provide valuable insights into public opinion that can be leveraged for decision-making purposes. Machine learning has become a popular approach for sentiment analysis (Alpna Patel et al.,2019). We are currently living in an age dominated by data, and Machine Learning is one of the most influential and powerful developments of our time (Anandhavalli Muniasamy et al.,2020). Machine learning is a valuable tool that can help turn data into knowledge. In the past 50 years, there has been an explosion of data, but this data is of little use unless we can analyze it and uncover hidden patterns. Machine learning techniques can be applied to complex and unstructured data to identify important hidden patterns and make sense of them (Arfan Shahzad et al.,2020).

## Impact of COVID-19 Outbreak on Higher Education in Pakistan

The COVID-19 outbreak first occurred in Wuhan, China in late December of 2019 and rapidly spread across the globe, ultimately being declared a pandemic by the World Health Organization on March 11th, 2020. The impact of this pandemic has been felt by countries both developed and developing, resulting in widespread closures of schools and other institutions as governments work to contain the virus. As a result, educational institutions around the world have been forced to transition from traditional face-to-face learning to online platforms, as noted in reference (Ashima Yadav et al.,2019). The scale of the impact on education has been immense, with more than 1.5 billion young people affected by the closure of schools and universities worldwide, as outlined by (Azoulay et al.,2021). The shift to online learning has presented numerous challenges for educators and students alike, including technological barriers, lack of access to reliable internet and devices, and difficulties in maintaining engagement and motivation without in-person interaction. As the pandemic continues to evolve, the future of education

remains uncertain, with many questions about the safety of returning to physical classrooms and the long-term impact on educational systems and outcomes. The pandemic (Cara Friedman et al.,2020) has disrupted traditional educational practices, forcing institutions to adopt alternative methods such as online learning to ensure that students continue to receive an education. It has also highlighted the need for increased investment in digital resources and infrastructure to support remote learning. On February 26, 2020, Pakistan recorded its first two cases of COVID-19, prompting the government to close all schools in the country on March 13, 2020 (Cara Friedman et al.,2020). The Higher Education Commission (HEC) then instructed higher education institutions to switch to online learning, suspend ongoing examinations, and offer consistent online support to students until the COVID-19 situation improved, as stated by (Costola, M et al.,2020). This response was aimed at minimizing the spread of the virus and ensuring that students continued to receive an education through alternative means, such as remote learning. According to data reported to the World Health Organization (Dr. G. S. N. Murthy et al.,2020) between January 3, 2020 and March 10, 2022, Pakistan has confirmed 1,516,150 cases of COVID-19 and 30,287 deaths. As of February 27, 2022, over 213 million doses of COVID-19 vaccines have been administered globally.

The COVID-19 pandemic has had far-reaching impacts on different sectors worldwide, including education, which has undergone a significant shift toward online learning in many countries (DR. Pravat Kumar Jena et al.,2020). Some countries were able to make a swift transition to online learning due to their existing e-learning infrastructure, as noted in a study by (Dr. Priyanka Harjule et al.,2020). However, developing countries like Pakistan face various challenges in Online Education, such as the lack of Internet availability in remote/rural/tribal areas, power outages, inadequate resources, and facilities, online examination difficulties, low computer literacy rates, unaffordable Internet data packages for students from low-income backgrounds, a shortage of electronic devices like laptops/smartphones, unsuitable learning environments, mental health concerns, difficulty comprehending online classes, and other related issues (Drus, Z et al.,2019). Despite following the example of developed countries by implementing online learning during the COVID-19 pandemic, Pakistan has struggled due to these challenges.

## Contribution

The contributions to this research illustrate:

- To classify evaluations as being in a positive, negative, or neutral tone, compile a sizable collection. This gave us important information about what students were thinking and feeling throughout the pandemic, which could utilize to enhance the effectiveness of our online courses.
- To analyze the sentiments stated in reviews of online courses, created a machine learning classifier-based system. This method might be utilized to increase the model's accuracy. This could be achieved by growing the dataset's size and contrasting it with current e-learning studies.
- Developed a sentiment analysis method based on machine learning that can glean information about students' opinions on online courses from their social media posts.

- This research offers assistance to the Higher Education Commission (HEC) of Pakistan, government officials, and policymakers to address and resolve these challenges and issues that students face in E-Learning classes. By working collaboratively, it is possible to develop targeted interventions and solutions that can improve the quality of E-Learning classes and ensure that higher education remains accessible and effective for all students, regardless of their circumstances.

### Paper Organization

The rest of the paper is organized as follows: Section.2 covers the background and relevant work; Section.3 covers research methods and framework; Section.4 covers results and discussion; and Section.5 contains the conclusion and future work.

## LITERATURE REVIEW

In this section, we have covered all the relevant research carried out on the topic of sentiment analysis utilizing machine learning. Additionally, we have addressed all of the methodologies and strategies that have been previously implemented. Eftekhar Hossain et al. (2020) was conducted the study using two datasets that already existed. The aim was to determine how well different machine learning algorithms can analyze the emotions expressed in tweets. The study provides detailed results on the accuracy achieved by various models, such as MNB, LR, SVM, RNN with LSTM, and Ensemble, in the two datasets. A research study was conducted by Erwin et al (2019) focusing on e-learning during the COVID-19 pandemic. The study analyzed the strengths, weaknesses, opportunities, and challenges (SWOC) of this form of education. One of the potential barriers identified in the study was the high cost of technological gadgets and internet plans, which could lead to losses for students. The authors also provided recommendations for addressing the challenges and problems encountered during e-learning in the context of the pandemic.

Ethem Can et al. (2020) to assess the capacity of Pakistani students for online and distance learning courses during the pandemic. The study discovered that numerous students experience financial difficulties that restrict their access to the internet, which ultimately affects the effectiveness of online learning. Hong Liang et al. (2017) conducted a cross-sectional study that utilized self-reported surveys to evaluate the efficacy of online courses for Pakistani students. The study acknowledges the advancements made in modern educational technology but proposes that further efforts are required before e-learning can entirely replace traditional classroom instruction. Jayaram A, S et al. (2020) the author conducted a thorough review of several popular deep-learning models used for sentiment analysis in their research. The study highlights the efficacy of these models in addressing sentiment analysis problems. Various deep learning techniques, including CNN, RNN, LSTM, GRU, SVM, Naïve Bayes, and DBN, were applied in the study. The most notable level of accuracy was attained using CNN+RNN, which achieved 93.2% accuracy on the IMDB dataset.

Khadija Alhumaid et al conducted a study by administering an online survey to students at Graphic Era Hill University in Dehradun, India. The survey aimed to evaluate the challenges associated with online education, such as internet connectivity and class participation. The main objective of the study was to gain insight into the viewpoints of

students regarding the possibility of online education in light of the difficult circumstances caused by the COVID-19 pandemic and subsequent lockdown measures. The study aimed to understand how students perceived the feasibility of online learning during these challenging times. Khadijah Mukhtar et al, the author investigates and explore the viewpoints of both teachers and students on the advantages, disadvantages, and suggestions for online examinations during the COVID-19 pandemic. The research aimed to gather insights from both stakeholders in the education system to facilitate better decision-making and the implementation of effective strategies for remote assessment during the pandemic. By examining the perspectives of teachers and students, the study sought to obtain a comprehensive understanding of the challenges associated with online exams and to provide recommendations to address these challenges. Erwin E. Rotas (2020) and his team conducted research using a qualitative research method to provide a description of the difficulties encountered by university students in the Philippines with regard to distance learning during the COVID-19 pandemic. The study aimed to explore the experiences of students in this challenging situation and to identify the various obstacles that they faced. By using qualitative research methods, the team was able to gather in-depth information about the students' perspectives and experiences, which could provide insights for educators, policymakers, and other stakeholders in the education system. The research findings could be used to inform the development of strategies and policies that would facilitate the effective delivery of distance learning in times of crisis.

Pranali et al. (2016) conducted a review aimed at exploring machine learning-based approaches to sentiment analysis. The research focused on identifying the key features of the techniques used in sentiment analysis, which is the process of determining the emotional tone of a piece of text. The study aimed to provide a comprehensive understanding of the machine learning-based approaches used in sentiment analysis and the benefits and limitations of these techniques. By examining the essential features of these techniques, the research aimed to provide insights into their applications and to identify the areas where they can be improved. The findings of this review could be useful in improving the effectiveness of sentiment analysis in various domains such as social media monitoring, market research, and customer experience management. Pravat Kumar Jena et al. (2020) presents a comprehensive analysis of the significant impacts of COVID-19 on the higher education system in India. The author obtained information from various sources, including websites, journals, and e-content, to provide an all-encompassing understanding of the situation. The research aims to provide insights into the challenges and opportunities presented by the pandemic in the higher education system in India. The study identifies the various issues faced by students, faculty members, and universities in response to the pandemic and the resulting lockdown measures. The research analyzes the changes in the delivery of education, including the transition to online and distance learning modes. The study also examines the impact of the pandemic on the accessibility and affordability of higher education, as well as the initiatives taken by the government and universities to address these challenges. Overall, the research aims to provide a comprehensive overview of the current state of higher education in India during the pandemic and to provide recommendations for potential strategies to mitigate the challenges faced by the sector. Dr. Pravat Kumar and Jena (2020) conducted a study primarily focused on the effects of the COVID-19 pandemic

on the Open and Distance Learning (ODL) system of the Indira Gandhi National Open University (IGNOU). The study aimed to identify the challenges and opportunities that emerged as a result of the pandemic, specifically related to the ODL system. By conducting this research, the team aimed to understand the impact of the pandemic on the delivery of education through distance learning and to discover the strategies adopted by IGNOU to tackle the difficulties and opportunities created by the pandemic. The study aimed to provide insights into the various measures implemented by IGNOU to ensure the continuity of education for their students and to analyze their effectiveness. The research findings could be useful in providing guidance to other distance learning institutions facing similar challenges due to the pandemic. Overall, the study aimed to provide a comprehensive overview of the impact of the pandemic on the ODL system of IGNOU and the strategies adopted by the university to mitigate the effects of the pandemic on distance education. Sukanta et al. (2020) highlights the challenges and opportunities that arose for stakeholders in higher education during the COVID-19 lockdown period. The author emphasizes the importance of conducting a comprehensive assessment of these challenges and opportunities, as well as the use of secondary data, to draw meaningful conclusions. The study aims to provide insights into the key issues faced by universities, students, faculty members, and other stakeholders in the education system in response to the pandemic. The research identifies the various challenges posed by the sudden shift to remote learning and online education, including technological limitations, lack of access to resources, and issues with pedagogy and student engagement.

Additionally, the study analyzes the opportunities presented by the pandemic, including the potential for increased access to education and the development of innovative models of learning. The author recommends that universities and policymakers conduct a thorough assessment of these challenges and opportunities, with a particular focus on using secondary data to inform decision-making. The research findings could be useful in developing strategies to mitigate the challenges faced by the education system and to take advantage of the opportunities presented by the pandemic. A group of researchers, including Eftekhar Hossain et al. (2020) introduced a technique that utilizes deep learning and BiLSTM to categorize restaurant customers' reviews as either positive or negative. They gathered a database of 8,435 reviews to assess their approach, and they utilized Google Colaboratory to develop deep learning applications. The results demonstrated that the BiLSTM strategy had the greatest precision, reaching 91.35% on the test dataset. Olivier Habimana et al. (2020) conducted a study that explored various deep learning techniques employed in different sentiment analysis tasks and assessed their performance on specific datasets. The study also examined the development patterns of deep learning models utilized in sentiment analysis tasks.

Olivier Habimana et al. (2020) authors contribute in tow folds. The first section provides an overview of traditional sentiment mining methods, while the second section explains various conventional techniques for sentiment analysis that rely on deep learning. They examined the strengths and limitations of sentiment analysis and proposed a framework for future studies. Mahona Joseph Paschal and co-researchers conducted a study that focused on e-learning courses offered by various African universities, particularly five universities during the Covid-19 outbreak. The research was guided by the Constructivist theory, and its aim was to comprehend the challenges experienced by e-learning

systems and the factors that enhanced their implementation during the Covid-19 pandemic. Mohammed Amin Almaiah et al. (2020) conducted a study aimed at providing valuable insights to policymakers, developers, and researchers about the key aspects of e-learning during the pandemic (2020). Dr. G. S. N. Murthy et al. (2020) conducted sentiment analysis on textual reviews using Long Short-Term Memory (LSTM) neural networks, known for their ability to handle large amounts of data. They collected data from various online sources, including TripAdvisor, Amazon, IMDB, and social media platforms such as Facebook and Twitter. Their study demonstrated an 85% accuracy rate for the LSTM deep learning approach and provides valuable insights for sentiment analysis on online reviews. Alpna Patel et al. (2020) conducted a study on sentiment analysis using various techniques for feature selection, sentiment classification, and deep learning. The IMBD dataset, comprising 50,000 reviews, was used to explore these methods. The researchers found that the use of recurrent neural networks (RNN) resulted in an accuracy rate of 87% for sentiment analysis. Qurat Tul Ain et al. (2017) conducted a study to identify the primary challenges that students face during online learning, specifically related to technology and internet connectivity.

The study aimed to understand the major disruptions students face while attending virtual classes and assess whether students have access to reliable technological tools and sufficient internet connectivity for online learning. Rajesh et al. (2020) conducted a study to evaluate the sentiment of an e-learning survey dataset. Their objective was to predict the most appropriate n-gram model that could extract machine learning features to achieve this. Ethem Can et al. (2020) developed a transfer learning-based sentiment analysis model in one language to apply it to sentiment analysis tasks in other languages, where the availability of data for analysis is limited. Jayaram et al. (2020) conducted a study to explore students' perspectives on the learning process in both traditional classroom and online teaching environments, specifically in the post-COVID situation. Khadija Alhumaid in contributed to examine the attitudes of teachers towards e-learning as an alternative to traditional classroom-based learning during the COVID-19 pandemic. The main objective of their research was to develop effective policies and strategies to promote e-learning by gaining insights from the teachers' perspectives. Rochyani Lestiyanawati et al. (2020) conducted a study to identify the strategies used and challenges faced by professors in implementing e-learning during the COVID-19 pandemic.

Their research aimed to understand the experiences of professors with online learning and to provide insights into effective strategies for future implementation. Seungwan Seo et al. (2020) conducted a comparative study on the different designs of deep learning classification models to determine their impact on sentiment classification models. To enhance e-learning environments, Parminder Kaur et al. (2020) conducted a study to explore the potential of neural networks. Their research demonstrated the successful implementation of these networks in improving e-learning systems. In their study, Qurat Tul Ain et al. (2017) applied deep learning models such as deep neural networks and convolutional neural networks to address various research challenges associated with sentiment analysis. They found that these models were effective in achieving more precise sentiment analysis results. Anandhavalli Muniasamy et al. (2020) conducted a study that provides an overview of how deep learning techniques can improve e-learning platforms by creating personalized learning experiences using prediction,

algorithms, and analysis. The study showed that deep learning can help reduce the costs of enhancing e-learning platforms by reusing existing resources. Hong Liang et al.(2017) applied traditional text feature extraction methods along with deep learning techniques in their study on diabetes patients from 301 hospitals. Mushtaq Hussain and his co-authors utilized machine learning algorithms and technological learning (TEL), also known as digital electronic education and design (DEEDS), to forecast potential difficulties that students may face when studying a digital design course. Abinash Tripathy et al. (2016) employed four machine learning models, including Naive Bayes (NB), Maximum Entropy (ME), Stochastic Gradient Descent (SGD), and Support Vector Machine (SVM), to categorize human sentiments. They evaluated the effectiveness, precision, F-measure, and accuracy of each approach in their research. Shilpa Singh Hanswal et al. (2020) gathered 500 tweets related to e-learning and implemented three machine learning algorithms - Naive Bayes, Support Vector Classifier, and Logistic Regression - to predict student dropout and completion rates in e-learning. The accuracy of each algorithm differed, with Naive Bayes achieving 61% accuracy, Logistic Regression achieving 70%, and Support Vector Classifier achieving 73.4% accuracy. The objective of this research is to improve the present research on E-learning during the COVID-19 pandemic by gathering a more comprehensive dataset from social media and leveraging Gaussian NB and Multinomial NB classifiers to enhance model accuracy in sentiment analysis. The study intends to identify the difficulties and problems faced by Pakistani higher education students during E-learning courses and analyze their attitudes towards E-learning in the context of the pandemic.

## Materials and Methods

In this section Proposed methodology is presented, the aim is to analyze the sentiment of Pakistani university students towards e-learning during the COVID-19 pandemic using a machine learning approach. We collected data from students through social media sites and used natural language processing techniques and machine learning algorithms to analyze the sentiment of students' responses. The study found that the overall sentiment towards e-learning was negative, with students appreciating the flexibility and convenience it offered. Students also faced challenges related to technical issues and a lack of interaction with teachers and peers. The passage presents a suggested framework that employs machine learning to forecast sentiment analysis in E-learning classes amidst the COVID-19 outbreak as shown in Fig. 1.

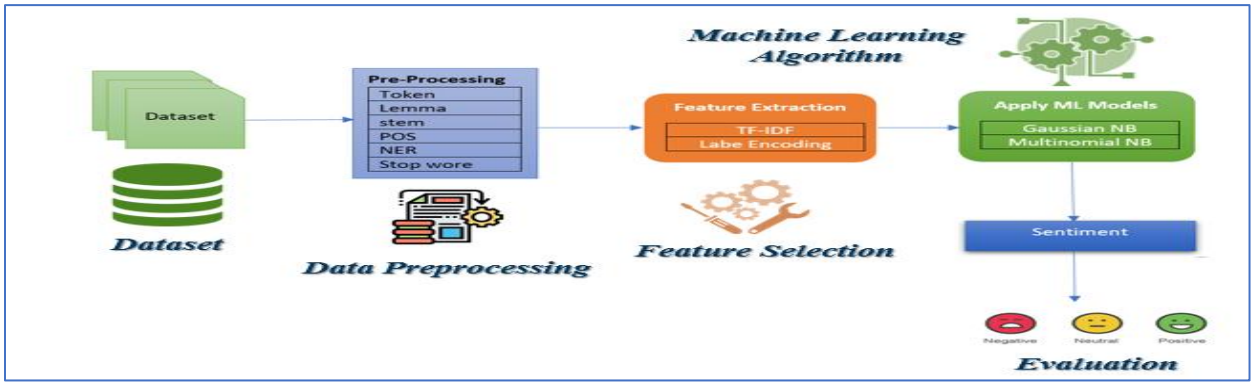


**Figure 1.**
**The architecture of our proposed system**

## Dataset and Challaenges

To conduct our experiment, we first accumulate reviews/comments from students across various social media platforms in Roman Urdu and English, and store the collected data in a designated location. During the COVID-19 pandemic, we spent 185 days gathering this dataset by scouring Facebook pages, YouTube channels, and Twitter accounts of multiple Universities and colleges in Pakistan to obtain student feedback and evaluations of E-learning courses. This dataset was subsequently utilized in our study for sentiment classification utilizing machine learning algorithms. We processed the data and utilized feature extraction approaches to create a feature vector suitable for training and testing the classifier models. We have collected data from many social media sites including Facebook, Twitter, YouTube, etc to get students' thoughts and comments. Roman Urdu and English words make up the data set that has been collected and are kept in a separate repository. This information is crucial to the success of our experiment. The challenges and issues that we identified on social media platforms like Facebook, Twitter, YouTube, etc. are listed in Table 1.

**Table 1.**
**Core Challenges faced by Pakistani University Students in E-Learning during COVID-19 Pandemic**

| S.NO | Challenges and Issues | S.No | Challenges and Issues |
|---|---|---|---|
| 1 | Internet availability | 2 | Technical Subject Understandability |
| 3 | Proper Resources and best Facilities | 4 | An improper platform for Online Classes |
| 5 | Online Examination issue due to poor internet connectivity | 6 | Untrained Students for Online Classes |
| 7 | Semester Fee Relaxation | 8 | Online Classes/Online Examination effect on Students Grades and CGPA |
| 9 | Computer Literacy | 10 | Online Classes quality and good education |
| 11 | Internet Data Packages i | 12 | Student's Lack of interest in the study |
| 13 | Poor internet connectivity issue in Remote/Rural/Backward areas of Pakistan | 14 | The heavy burden of Classes/Assignments/Quiz etc. |
| 15 | PC/Laptop/Smartphone availability | 16 | Lack of Coordination between Students and Teacher |
| 17 | Load shedding issue in Remote/Rural/ Backward areas of Pakistan | 18 | Students' Lack of interest in Online Classes |
| 19 | Proper Environment issue for Online Classes | 20 | Financial Issue |
| 21 | Internet Device | 22 | Highly-paid Semester Fee |
| 23 | Online Classes Timing | 24 | Proper Training |
| 25 | Technical | 26 | Unscheduled Classes |
| 27 | Online Classes sound Quality | 28 | Students Self-motivation |
| 29 | Mental Health Issue | 30 | Technology issue |

| | | | |
|---|---|---|---|
| 31 | Untrained Teacher/Instructor for Online Classes | 32 | Online Classes Disturbance issue |
| 33 | Online Classes Understandability | 34 | interruption |
| 35 | Highly-paid Internet Data Package | 36 | Majority of Faculty is unfamiliar with how to use the Online System |
| 37 | Proper Training issue for Online Classes | 38 | Online Classes Concentration |
| 39 | Teacher lack of interest in Online Classes | 40 | Labs and Practical |

## Data Pre-Processing

Data pre-processing is the initial step in any data analysis project that involves cleaning and transforming raw data to make it usable for further analysis. Data cleansing, data transformation, data normalization, data integration, data reduction, and data discretization are just a few of the procedures that are used. Data pre-major processing's goal is to make sure the data is reliable, accurate, and consistent so that it can be evaluated properly. Because the quality of the data used for training and testing a model determines how accurate it will be, data pre-processing is essential in machine learning and data mining projects. Once the dataset is obtained, it undergoes pre-processing to eliminate any outliers or duplicate records that may impede its effectiveness. This procedure, known as data wrangling, encompasses cleaning the data by identifying and rectifying any erroneous entries in a table or database. This involves detecting incomplete or irrelevant data and taking corrective steps such as substituting, modifying, or eliminating them. Text preprocessing includes several operations such as Tokenization, Parts of Speech tagging, Data Cleaning, stop word Removal, and Stemming.

## Feature Extraction

Feature extraction is the process of converting raw data into numerical features that can be utilized in processing while still retaining the information present in the original dataset. This strategy produces superior results than directly implementing machine learning algorithms to raw data. The process of selecting and transforming raw data into a set of pertinent features that may be utilized to portray the data in a more useful and effective way is known as feature extraction. Feature extraction is essential to machine learning and data analysis because it lowers the dimensionality of data, enhances model performance, and makes data visualization easier. Techniques including dimensionality reduction, scaling, normalization, and transformation are employed during the feature extraction process to change the raw data into a format that machine learning algorithms can quickly examine. Natural language processing, image, and speech recognition, as well as other applications that deal with high-dimensional data, all benefit greatly from feature extraction.

## TF-IDF Model

TF-IDF is an essential feature extraction process for evaluating the characteristics of a review. Before the TF-IDF algorithm is applied, pre-processing is conducted. After pre-processing, sentences are tokenized instead of individual words. A weight value is then assigned, and word frequency is calculated. During this step, the Term Frequency (TF) is computed using the formula in Eq.1.

$$TF(t) = \frac{Number\ of\ times\ term\ t\ appear\ in\ a\ document}{Total\ number\ of\ terms\ in\ the\ document} \quad (1)$$

Following that, a table is established to document the frequency of each word in each sentence. Then, the inverse document frequency (IDF) is determined using the formula in Eq. 2.

$$IDF(t) = \log e \frac{Total\ number\ of\ documents}{Number\ of\ documnets\ with\ term\ t\ in\ it} \quad (2)$$

The TF-IDF score is determined by multiplying the values of equation (1) and equation (2). In order to determine the threshold value, the average score of all words is computed during these steps. Any words with a score higher than the threshold value are selected. The overall process of the TF-IDF method is illustrated in Figure 1. Prior to determining the TF-IDF scores, we employed the stemming process to obtain the root of each word by removing prefixes and suffixes. We utilized two classifier models, Gaussian naïve Bayes (GNB) and multinomial naive Bayes (MNB).

## Naïve Bayes Classifier

In our study, we utilize the Naive Bayes Classifier to forecast the sentiment of higher education students regarding E-learning classes during the COVID-19 pandemic. Bayes' theorem is a mathematical formula used to calculate the probability of a hypothesis based on previous knowledge, which is determined by conditional probability. The Naive Bayes algorithm is a computationally efficient technique used in text analysis that assigns equal importance to all words in order to capture every single phrase in the text. The Naive Bayes algorithm applies Bayes' theorem, which assumes that the features are independent of one another, implying that the probability of one feature is not affected by another feature. The probability model is expressed in Eq.3 The formula for Bayes' theorem is as follows:

$$P(A|B) = (P(B|A) * P(A)) / P(B) \quad (3)$$

## Gaussian Naïve Bayes Classifier

A probabilistic approach used in machine learning for classification tasks is called Gaussian Naive Bayes (GNB). It is based on the Bayes theorem and the presumption that given the class, the input data's properties are conditionally independent of one another. The Bayes theorem, which states that the probability of a hypothesis (class) given the data (features), is proportional to the probability of the data given the hypothesis multiplied by the prior probability of the hypothesis, is used in GNB to calculate the probability of an instance belonging to a specific class. The procedure determines the mean and variance of each feature for each class under the assumption that the features have a Gaussian (normal) distribution. By multiplying these parameters together, the conditional probability of each feature given the class is calculated, and the result is the chance that the instance belongs to that class. GNB is a quick and straightforward method that performs well with high-dimensional data and is frequently used as a starting point for more intricate ones. However, if the requirement of conditional independence is not met or if the distribution of the features deviates sufficiently from a Gaussian

distribution, its performance may deteriorate. If the predictors are continuous rather than discrete, the Gaussian model expects the continuous values to be drawn from the Gaussian distribution because it is based on the assumption that features follow a normal distribution

### Multinomial Naïve Bayes Classifier

A probabilistic method called the (MNB) classifier is used in ml for classification tasks, particularly text categorization. It is a special case of the Naive Bayes method where the features are the counts or frequencies of the words or terms in the text data and it is specifically made for handling discrete data. The MNB classifier is based on the presumption that, given the class label, each feature is conditionally independent of every other feature. Using the training data and the product rule, it determines the joint probability of all the features given the class label. It then estimates the chance of each feature occurring in a document for each class. The predicted class label for the dataset is given to the class with the highest joint probability. Whether the input data is in the form of a term-frequency matrix or a bag of words, the MNB classifier is very helpful for text classification tasks like spam filtering, sentiment analysis, and topic categorization. It has a lot of features, is quick and easy to use, and can manage big volumes of data. The Multinomial Naive Bayes classifier is applied to multinomial distribution-following data. This classifier is very useful for document classification jobs, which include categorizing a document into a particular category, such as Sports, Politics, Education, etc. The classifier's predictors are based on the frequency of terms inside the document.

## EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we discuss the outcomes of our proposed approach to overcome the challenges that higher-education students in Pakistan faced with E-learning during the COVID-19 pandemic, as well as how the machine-learning models play an effective role in terms of different evaluation parameters.

### Implementation Details

The studies utilized two Python modules and packages, namely sci-kit-learn and Seaborn. The experiments were conducted on a machine equipped with an Intel Core i5 processor and 8 GB of RAM. The studies were executed using Python 3.9 on the Spider IDE.

### Data Visualization

Data visualization is essential for comprehending and interpreting the data when using machine learning methods. The data can be visualized to reveal patterns and trends, outliers, and insights that can be used for feature selection, model selection, and performance evaluation. Scattered plots, heat maps, box plots, and histograms are a few examples of data visualization tools that can be used to show the distribution and connections between different variables. The performance of deep learning models can also be tracked using tools like Tensor Board, which allows users to visually track the training process. a large number of attributes.

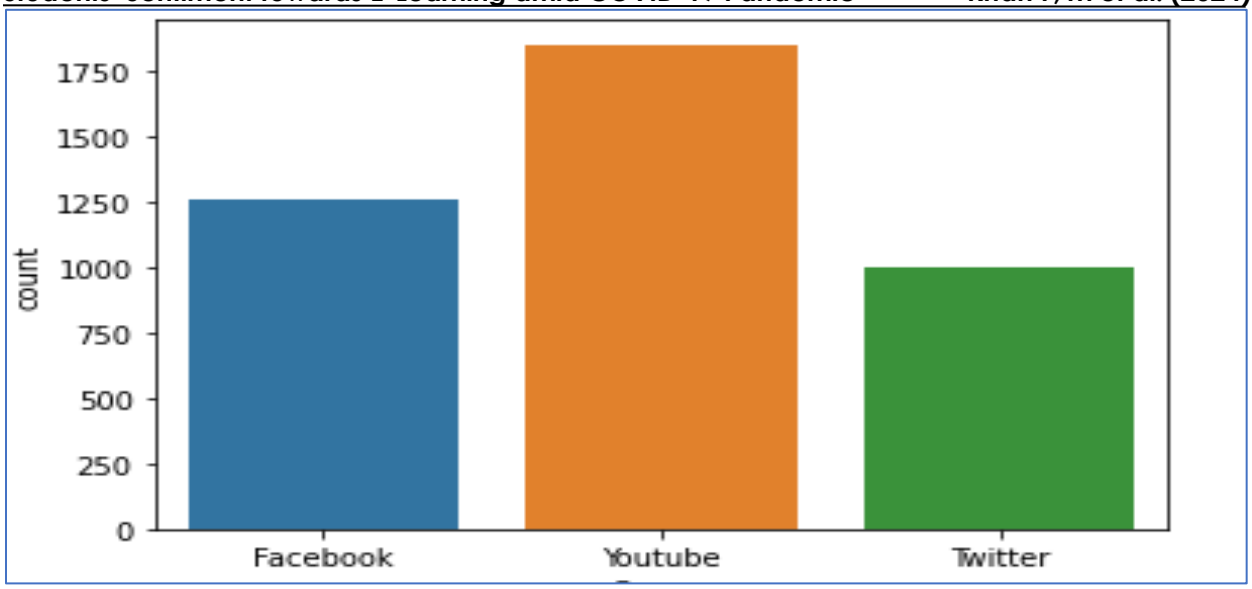


**Figure 2.**
**Dataset Source**

Fig. 2 displays the source of the dataset, depicting those 1850 data points collected from YouTube, 1257 from Facebook, and 999 from Twitter.

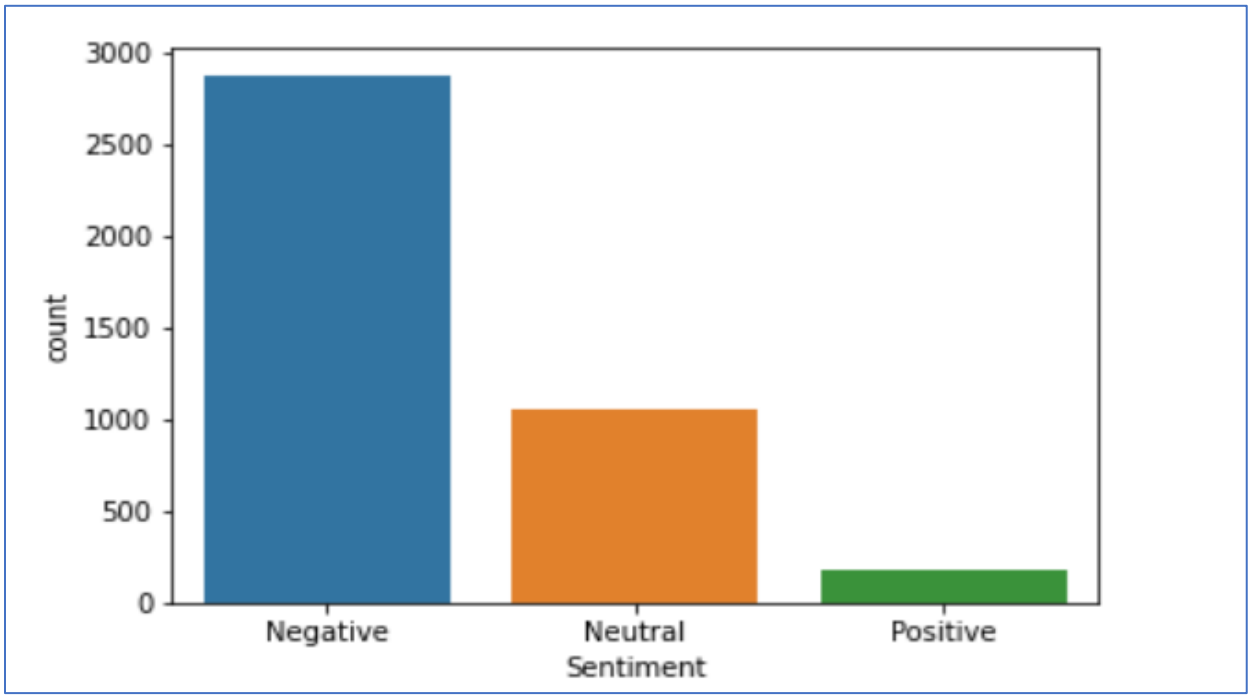


**Figure 3.**
**Actual Sentiment**

Fig. 3 displays the sentiment analysis of E-learning classes during the COVID-19 pandemic. The sentiment of students toward online classes is classified into negative, neutral, and positive categories. The negative sentiment has 2880 occurrences, the neutral sentiment has 1054 occurrences, and the positive sentiment has 172 occurrences. These sentiments are the target classes for our proposed model.

## Confusion Matrix of Gaussian Naïve Bayes Classifier

A confusion matrix heatmap is a graphical representation of a confusion matrix that employs color-coded cells to show the distribution of occurrences of various classes in a classification model that were successfully and wrongly categorized. The heatmap shows where the majority of classification errors are occurring and gives an easy-to-read assessment of the model's performance as shown in Fig.4.

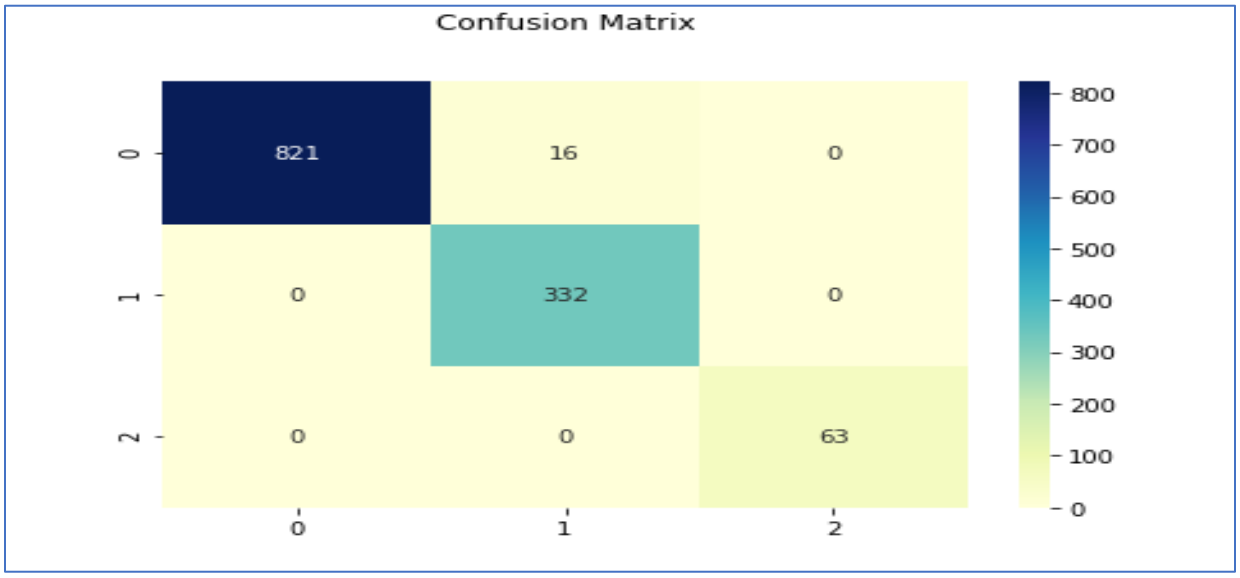


**Figure 4.**
**Confusin Matrix of Gaussian Naïve Bayes**

## Prediction Results of Guassian Naïve Bayes

The illustration in Fig. 5 displays the outcomes of our anticipated results using the Gaussian NB model that we suggested.

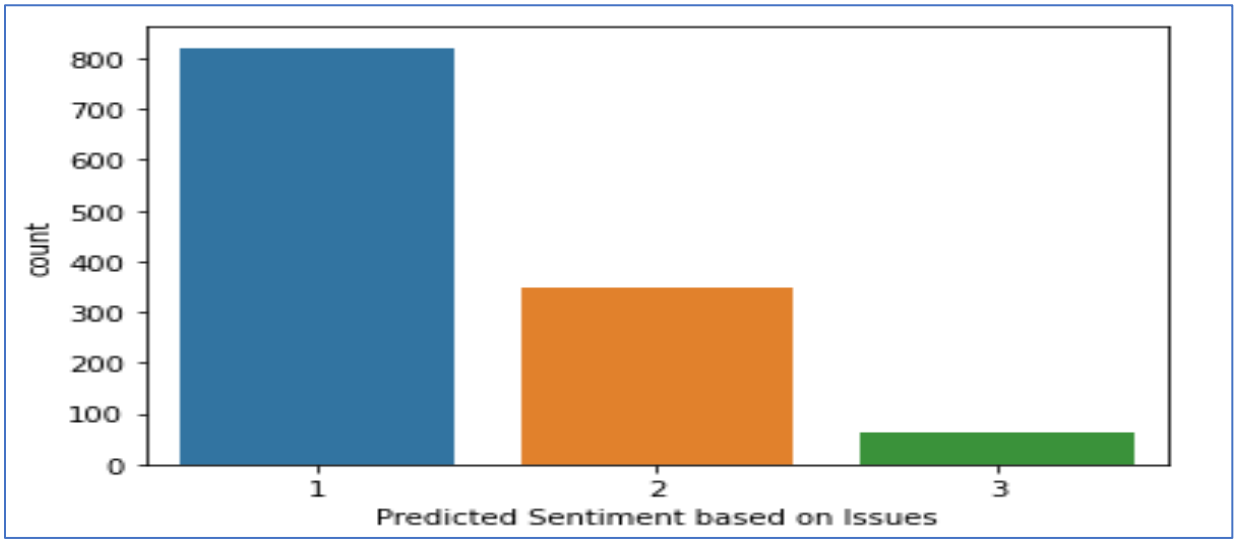


**Figure 5.**
**Prediction sentiment based on Issues**

## Confusion Matrix of Multinomial Naïve Bayes Classifier

The following confusion matrix pertains to the classification of the Multinomial NB model, which is displayed in Fig 6. It helps in identifying the performance of the Multinomial NB model.

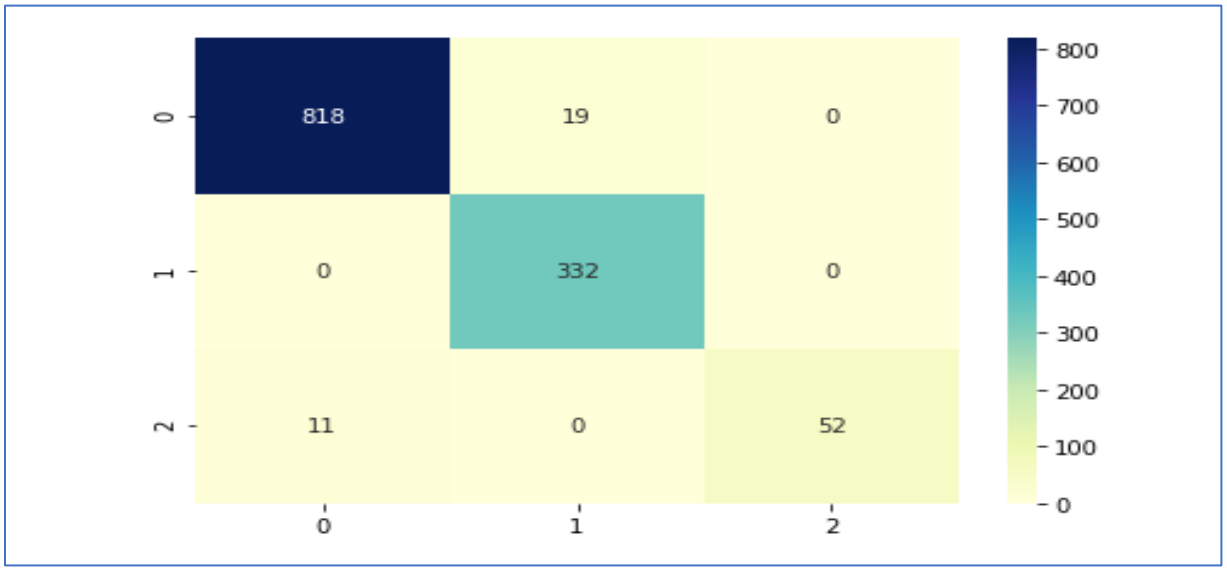


**Figure 6.**
**Confusion Matrix of Multinomial Naïve Bayes**

## Prediction Result of Multinomial Naïve Bayes

We have made a second prediction using the Multinomial NB model. Fig. 7 below demonstrates the classification results and predicted outcomes.

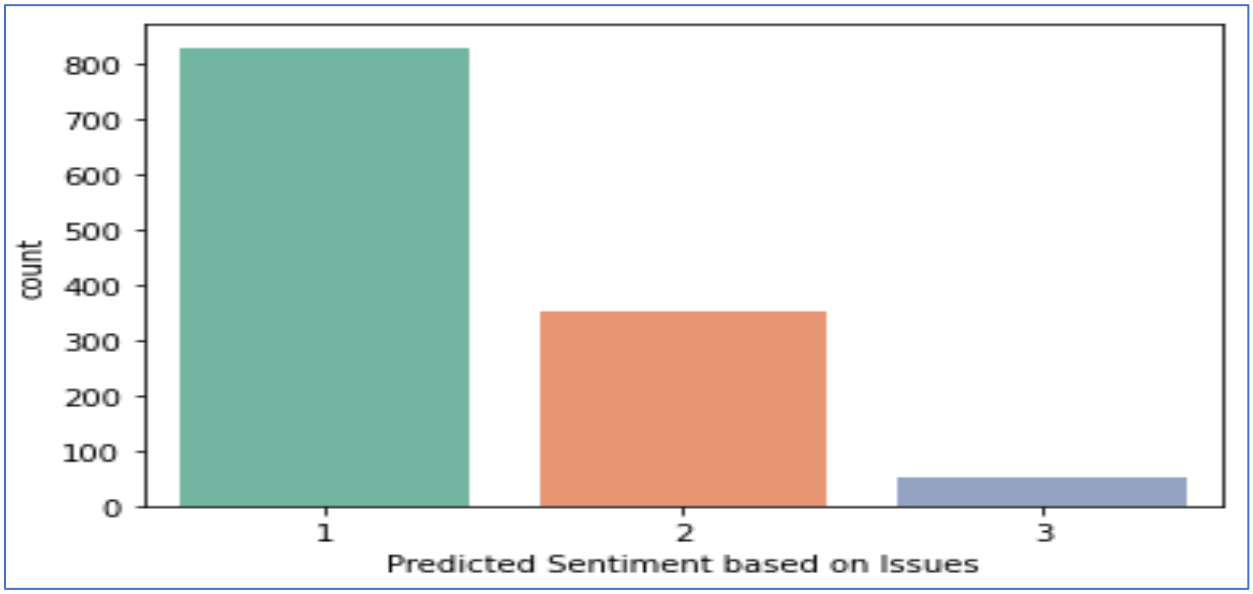


**Figure 7.**
**Prediction sentiment result**

In Figure 7 displays the number of negative, neutral, and positive outcomes, which are 829, 351, and 52, respectively. This represents a 30% validation test result on the real dataset. In comparison to the visualization of our actual dataset, our model's accuracy in terms of sentiments is 98% accurate.

**Table 2.**
**Strategic Recommendation/Suggestions**

| |
|---|
| 1. Pakistan must immediately adopt 5G and Metaverse Technology to improve online education through the provision of faster, more reliable internet connectivity. Online course materials and virtual classrooms may be more convenient for students. |
| 2. Make sure that students who don't have computers and the internet can use them. Facilities such as public libraries, community centers, and internet cafés can be used.<br>3. The government should launch initiatives to upgrade the internet infrastructure in remote regions.<br>4. Students who have limited access to power should be encouraged to study during the day, when electricity is more reliable, and alternate sources of light, such as<br>5. Solar lights, should be made available throughout the school/Colleges/Universities.<br>6. Student loans should be made available to students with low incomes so that they can purchase devices.<br>7. Provide students, teachers, and parents with opportunities to increase their digital literacy abilities through the implementation of initiatives.<br>8. students should be encouraged to join online communities and forums where they can connect with their peers and teachers. Isolation and a lack of social interaction pose a significant problem for pupils.<br>9. Offering counseling and support services to students who are experiencing stress, anxiety, and other mental health difficulties related to online learning is the eighth recommendation for addressing mental health concerns.<br>10. It is important to provide instructors with chances for professional development to assist them in properly delivering classroom education online.<br>11. It is important to offer training programs for parents to assist them in comprehending the significance and advantages of online education.<br>12. Establishing criteria for online learning and monitoring the quality of education that is being given to verify that it satisfies these standards is the first step in addressing the issue of inconsistent educational quality. |

# RESULTS

The proposed methods are evaluated on the accuracy, precision, recall and f1 score, the outcomes of the proposed models are reflected in the table below. The comprehensive classification is shown in Table 3.

**Table 3.**
**The classification result of classifiers**

| ML Classifier | | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| Gaussian Naïve Bayes | Macro Avg | 98% | 0.99 | 0.99 | 0.99 |
| | Weighted Avg | 99% | 0.99 | 0.99 | 0.99 |
| Multinomial Naïve Bayes | Macro Avg | 97% | 0.95 | 0.93 | 0.95 |
| | Weighted Avg | 98% | 0.98 | 0.98 | 0.98 |

# CONCLUSIONS AND FUTURE WORK

The objective of this research is to examine the impact of the COVID-19 pandemic on E-learning courses for higher education students in Pakistan. The findings indicate that students are encountering various obstacles in online classes, including issues with electricity and internet connectivity, a lack of resources and amenities struggles with comprehending online content, mental health concerns, and more. The Gaussian NB and Multinomial NB classification algorithms were employed for the machine learning model, and a confusion matrix was generated to assess its performance. The study achieved high accuracy compared to other existing research studies, by increasing the dataset size of E-learning courses during the pandemic. The conclusions of this study can assist the Higher Education Commission, the government, and policymakers in addressing these challenges and enhancing the E-learning experience for students during and after the pandemic. It is recommended that further research be conducted to establish a deep learning predictive model for forecasting the future impact of COVID-19 on student's education.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Dataset will be provided on demand.

**Conflicts of Interest:** The authors declare no conflict of interest.

# DECLARATIONS

**Acknowledgement:** We appreciate the generous support from all the supervisors and their different affiliations.
**Funding:** No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
**Availability of data and material:** In the approach, the data sources for the variables are stated.
**Authors' contributions:** Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
**Consent to Participate:** Yes
**Consent for publication and Ethical approval:** Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their
consent.

# REFERENCES

Abinash Tripathy, Ankit Agrawal and Santanu Kumar Rath, "Classification of sentiment reviews using n-gram machine learning approach", *http://www.elsevier.com/locate/eswa, http://dx.doi.org/10.1016/j.eswa.2016.03.028*, 24 March 2016.

Adnan, M., & Anwar, K. (2020). Online Learning amid the COVID-19 Pandemic: Students' Perspectives. *Online Submission*, 2(1), 45-51.

Alpna Patel and Arvind Kumar Tiwari, "Sentiment Analysis by using Recurrent Neural Network", *2nd INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING AND SOFTWARE ENGINEERING*, https://ssrn.com/abstract=3349572, 2019.

Anandhavalli Muniasamy and Areej Alasiry, "Deep Learning: The Impact on Future eLearning", *https://doi.org/10.3991/ijet.v15i01.11435, http://www.i-jet.org/, 2020.*

Arfan Shahzad, Rohail Hassan, Adejare Yusuf Aremu, Arsalan Hussain and Rab Nawaz Lodhi, "Effects of COVID-19 in E - learning on higher education institution students: the group comparison between male and female", *https://doi.org/10.997/s11135-020-01028-z , © Springer Nature* B.V. 2020.

Ashima Yadav and Dinesh Kumar Vishwakarma, "Sentiment analysis using deep learning architectures: a review", *https://doi.org/10.997/s10462-019-09794-5, © Springer Nature B.V,* 02 December 2019.

Azoulay, UNESCO Director-General Audrey. "Global Education Coalition." UNESCO, *UNESCO, 7 June 2021, en.unesco.org/covid19/educationresponse/globalcoalition.*

Cara Friedman, "Students' Major Online Learning Challenges amid the COVID-19 Pandemic".

Costola, M., Nofer, M., Hinz, O., & Pelizzon, L. (2020). Machine learning sentiment analysis, Covid-19 news and stock market reactions.

Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, Mounika Bagadi and Mounika Belusonti, "Text based Sentiment Analysis using LSTM", *International Journal of Engineering Research & Technology (IJERT),* http:// www.ijert.org, May-2020.

DR. Pravat Kumar Jena, "Challenges and Opportunities created by Covid-19 for ODL: A case study of IGNOU", http://www.ijirmf.com/, I*NTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD*, Publication Date: 31/05/2020.

DR. Pravat Kumar Jena, "Impact of Covid-19 on higher education in India", https://www.researchgate.net/publication/342691816_Impact_of_Covid-19_on_Higher_Education_in_India? esc=publicationCoverPdf , DOI: 10.31235/osf.io/jg8fr, *International Journal of Advanced Education and Research*, July 2020.

Dr. Priyanka Harjule, Astha Gurjar, Harshita Seth and Priya Thakur, "Text Classification on Twitter Data", *3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE-2020)*, 07-08 February 2020.

Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: systematic literature review. *Procedia Computer Science*, 161, 707-714.

Eftekhar Hossain, Omar Sharif, Mohammed Moshiul Hoque and Iqbal H. Sarker, "SentiLSTM: A Deep Learning Approach for Sentiment Analysis of Restaurant Reviews", 2020.

Erwin E. Rotas and Michael B. Cahapay, "Difficulties in Remote Learning: Voices of Philippine University Students in the Wake of COVID-19 Crisis", *https://www.researchgate.net/publication/347885314*, December 2020.

Ethem Can and Aysu Ezen-Can, "Multilingual Sentiment Analysis: An RNN-Based Framework for Limited Data", *SAS Institute Inc, SAS #SASGF ® GLOBAL FORUM 2020.*

Hong Liang, Xiao Sun, Yunlei Sun and Yuan Gao, "Text feature extraction based on deep learning: a review", http://creativecommons.org/licenses/by/4.0/, *EURASIP Journal on Wireless Communications and Networking 2017*.

JAYARAM A S, "Student Survey Regarding On-Line Classes During and Post Covid-19 Period", *ICONIC RESEARCH AND ENGINEERING JOURNALS, © JUL 2020.*

Khadija Alhumaid, Sana Ali, Anbreen Waheed, Erum Zahid, Mohammed Habes, "COVID-19 &Elearning: Perceptions &Attitudes of Teachers Towards ELearning Acceptancein The Developing Countries", *DOI: 10.5281/zenodo.4060121, September 30, 2020.*

Khadijah Mukhtar, Kainat Javed, Mahwish Arooj and Ahsan Sethi, "Advantages, Limitations and Recommendations for online learning during COVID-19 pandemic era", *https://doi.org/10.12669/pjms.36.COVID19-S4.2785, 2020.*

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, 1-57.

Mahona Joseph Paschal and Demetria Gerold Mkulu, "Online Classes during COVID-19 Pandemic in Higher Learning Institutions in Africa", *http://dx.doi.org/10.22158/grhe.v3n3p1, Online Published: August 17, 2020.*

Mahyoob, M. (2020). Challenges of e-Learning during the COVID-19 Pandemic Experienced by EFL Learners. *Arab World English Journal (AWEJ), 11(4)*.

Mitra, A. (2020). Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset). *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, 2(03), 145-152.

Mohammed Amin Almaiah, Ahmad Al-Khasawneh and Ahmad Althunibat, "Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic", https://doi.org/10.997/s10639-020-10219-y, 22 May 2020.

Muhammad Adnan and Kainat Anwar, "Online learning amid the COVID-19 pandemic: Students' perspectives", http://www.doi.org/10.33902/JPSP.%202020261309, *Journal of Pedagogical Sociology and Psychology, 2020.*

Mushtaq Hussain, Wenhao Zhu, Wu Zhang, Syed Muhammad Raza Abidi and Sadaqat Ali, "Using machine learning to predict student difficulties from learning session data", https://doi.org/10.997/s10462-018-9620-8, *Published online: 10 February 2018.*

Nadeak, B. (2020). The effectiveness of distance learning using social media during the pandemic period of covid-19: A case in universitas kristen Indonesia. *International Journal of Advanced Science and Technology*, 29(7), 1764-1772.

Olivier HABIMANA, Yuhua LI, Ruixuan LI, Xiwu GU and Ge YU, "Sentiment analysis using deep learning approaches: an overview", https://doi.org/10.997/S11432-018-9941-6, *c Science China Press and Springer-Verlag GmbH Germany, January 2020.*

P.Rajesh and G.Suseendran, "Prediction of N-Gram Language Models Using Sentiment Analysis on E-Learning Reviews", *International Conference on Intelligent Engineering and Management (ICIEM)*, DOI: 10.1109/ICIEM48762.2020.9160260 June 2020.

Parminder Kaur, Kiranjit Kaur and Gurdeepak Singh, "Improving E-Learning with Neural Networks", *International Journal of Computing & Business Research, 2020.*

Pranali Borele and Dilipkumar A. Borikar, "An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques", http://www.iosrjournals.org/, Apr. 2016.

QASIM MEHMOOD, USMAN JAVED SAHI, RANA RAKHSHAN AFTAB , RANA KHURRAM AFTAB, FARHAT IJAZ and AVAIS AHMAD," ONLINE CLASSES IN DIFFERENT INSTITUTES OF PAKISTAN DURING CORONAVIRUS PANDEMIC; MERITS AND DEMERITS", http://www.ppmj.org.pk/, Jun. 2020.

Qurat Tul Ain, Mubashir Ali, Amna Riaz, Amna Noureen, Muhammad Kamran, Babar Hayat and A. Rehman, "Sentiment Analysis Using Deep Learning Techniques: A Review", *International Journal of Advanced Computer Science and Applications,* http://www.ijacsa.thesai.org/, 2017.

Rahman, A., & Hossen, M. S. (2019, September). Sentiment analysis on movie review data using machine learning approach. *In 2019 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-4). IEEE.

Raj Kishor Bisht, Sanjay Jasola and Ila Pant Bisht, "Acceptability and challenges of online higher education in the era of COVID-19: a study of students' perspective", https://doi.org/10.1108/AEDS-05-2020-0119, 14 August 2020.

Rashid, S., & Yadav, S. S. (2020). Impact of Covid-19 pandemic on higher education and research. *Indian Journal of Human Development, 14(2), 340-343.*

Rochyani Lestiyanawati and Arif Widyantoro, "Strategies and Problems Faced by Indonesian Teachers in Conducting ELearning System During COVID-19 Outbreak", *Journal of Culture, Literature, Linguistics and English Teaching, May 2020.*

SEUNGWAN SEO, CZANGYEOB KIM, HAEDONG KIM, KYOUNGHYUN MO AND PILSUNG KANG, "Comparative Study of Deep Learning-Based Sentiment Classification", http://creativecommons.org/licenses/by/4.0/, *publication January 1, 2020.*

Shilpa Singh Hanswal, Pareek, Astha & Vyas, Geetika & Sharma, Amita. (2020). Sentiment Analysis on E-Learning Using Machine Learning Classifiers in Python

Shivangi Dhawan, "Online Learning: A Panacea in the Time of COVID-19 Crisis", http://dx.doi.org/10.1177/0047239520934018 , *Journal of Educational Technology Systems, 2020.*

Sukanta Kumar Baral, "E-learning Challenges during COVID-19 Lockdown Period: A Revisited insight for Higher Education", http://www.crdeepjournal.org/, *Published: 18-12-2020.*

Ullah, A., Ashraf, M., Ashraf, S., & Ahmed, S. (2021). Challenges of online learning during the COVID-19 pandemic encountered by students in Pakistan. *Journal of Pedagogical Sociology and Psychology*, 3(1), 36-44. https://doi.org/10.33902/JPSP.2021167264

Wenling Li, Bo Jin and Yu Quan, "Review of Research on Text Sentiment Analysis Based on Deep Learning",https://doi.org/10.4236/oalib.1106174,http://creativecommons.org/licenses/by/4.0/, March 9, 2020.

WHO, WHO. "Pakistan: WHO Coronavirus Disease (COVID-19) Dashboard With Vaccination Data." World Health Organization, *World Health Organization, 2 July 2021, covid19.who.int/region/emro/country/pk.*

Zaidi, Saba; Salah, Saman (2020): Impact of Coronavirus on Higher Education in Pakistan.docx. Advance. *Preprint. https://doi.org/10.31124/advance.12655871.v1*