Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 233 (2024) 590–596

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on Innovative Data Communication Technologies and Application (ICIDCA 2024)

Student's Interest and Opinion Towards Online Education

Hemanth Sravan Kumar Veguru[a], Naren J[b], and Yasasree Singam[c] *

[a]*Department of Information Science and Technology, Missouri University of Science and Technology, USA, 65401, hemanthsravan1999@gmail.com*
[b]*Assistant Professor, Alliance University, Bangalore, 562106, India, naren.jeeva3@gmail.com*
[c]*Department of Data Science, Michigan State University, USA, 48824, singam.yashu11@gmail.com*

Abstract

The paper presents a study of the interest and opinions of students regarding online learning using a machine learning approach, as well as the evolution of different e-learning platforms based on education following the pandemic period. The study uses student data from a questionnaire-based survey of college students to improve the learning environment from the perspective of the students. The survey and questionnaire were designed with students' needs, requirements, and preferred level of quality for online learning. Survey data was subjected to data analysis and classification to gain a deeper understanding of the learner in a virtual learning environment. Through data preparation, analysis, visualization, and machine learning algorithm accuracy, machine learning classification algorithms and analysis are used to examine the collected data. The research study will illustrate the expectations and enhancement of online learning education patterns according to the requirements of students. With a 93% accuracy evaluation, the Random Forest algorithm has the best accuracy among the several classification algorithms.

Keywords: Online Education; Student's Perspective; Virtual Learning Environment; Machine Learning; Classification Analysis;

1877-0509 © 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 5th International Conference on Innovative Data Communication Technologies and Application
10.1016/j.procs.2024.03.248

1. Introduction

Online education has become increasingly popular in recent years, and the COVID-19 pandemic has accelerated this trend. The internet and technological advances have made it possible for students worldwide to access more interactive online learning environments and educational content. Several factors, including the need for flexibility, the desire to learn from professionals worldwide, and the growing expense of traditional education, have led to a noticeable increase in the demand for online education since the pandemic began. Along with its advantages, there are some noticeable problems associated with online learning such as students may feel isolated and miss out on social interaction and communication with teachers and classmates which is a primary key factor to be considered in comparison with traditional face-to-face learning and students might not be able to utilize laboratories and technical infrastructure facilities provided by educational institutions. Students may experience technical challenges like internet connectivity issues and software glitches**.** Students could therefore struggle to maintain their motivation in a virtual learning environment. The growing prevalence of online education has necessitated a deeper understanding of student perceptions and experiences in this learning environment. Thus, a mechanism to predict students' opinions and interest in the medium needs to be developed to improve the resources and current practices related to virtual learning.

The research has been premised on the development of a machine learning classification analysis and the evaluation of student data concerning their opinions and interest in online learning. Machine learning (ML) is a powerful method for examining and identifying patterns and trends in large datasets that would be challenging to identify using traditional methods of statistical analysis in the context of education. Machine learning can identify trends that present a learning obstacle for students or identify areas where students are dropping out of classes and could improve the online learning experience for all students. Numerous student outcomes, including academic success, dropout rates, and engagement levels, can be predicted using machine learning. Previous research has demonstrated that student perspectives on online learning can vary considerably as planning, executing, and assessing online learning is necessary to reduce issues and optimize the accomplishment of learning goals [12-15]. Some students find online learning to be a positive and engaging experience, while others find it to be challenging or isolating [1]. The dataset comprises student data about their enthusiasm and preferred methods of teaching in online learning. It includes columns with information on attendance percentage, overall rating, least-engaging aspect, preferred teaching approach, teaching assistance, learning resources allocation, percentage of efforts, satisfaction with learning, and excitement to attend [1-4][7-10]. The data will be utilized for the evaluation of online learning practices and to enhance teaching pedagogy. E-learning platforms can use such evaluations to better understand the perspectives and personal experiences of students in this kind of learning environment and to develop more effective teaching and learning strategies.

2. Literature Review

Gratiela Dana Boca [1] presented a classification analysis based on surveys used to determine university students' attitudes and behaviors about online learning during the pandemic. The four sections of the questionnaire were designed to determine the needs, preferences for quality in online education, online platform usage knowledge, and characteristics of the students. The study found that students generally prefer a hybrid learning model that combines online and in-person instruction, highlighting the need for a balanced approach that caters to diverse learning preferences.

Libby V. Morris et al. [2] suggested a multiple regression data analysis model to evaluate the predictive ability of student participation measures on attainment. Thirteen sessions, spread throughout three undergraduate general education classes, were used to perform the study. Using student access computer logs and student behaviors (defined as frequency of participation and duration of engagement) recorded for eight variables, data analysis was done on the records of 354 students. The study concluded that consistent participation and the duration of engagement are crucial for successful online learning, emphasizing the importance of active involvement in the learning process.

Qiu, Zhang, Sheng, et al. [3] introduced a behavior classification-based e-learning performance (BCEP) prediction system. To determine the category feature values for each type of behavior, it first chooses the characteristics of e-learning behaviors, creates a machine learning-based learning performance predictor, and then uses feature fusion with behavior data in line with the behavior classification model. The process-behavior classification (PBC) model, an online behavior classification model predicated on the e-learning process, was also proposed by the study. According to the study, the PBC model predicts learning performance better than traditional classification methods, and the BCEP prediction framework has a higher prediction impact. The study demonstrated the effectiveness of behavioral data in predicting student performance.

Jawad Khurram et al. [4] suggested integrating the data-balancing technique with a random forest classifier to gauge student interest and estimate academic growth. The Open University Learning Analytics Dataset (OULAD) was used in the study. Data from six different time periods, including engagements and assessment scores, were recorded to generate student profiles. The study highlighted the importance of data balancing in improving the accuracy of student engagement assessment models.

Khawlah Altuwairqi et al. [5] proposed an automatic multimodal technique that analyzes three modalities to represent students' behaviors: emotions from facial expressions, keystrokes on the keyboard, and mouse movements to assess student engagement levels in real time. The proposed multimodal technique consisted of three steps: feature extraction, data gathering, and engagement recognition. The authors validated the proposed multimodal strategy through three main experiments: single, dual, and multimodal research modalities in new engagement datasets. The method using many modes produced the highest accuracy result (95.23%). The study demonstrated the feasibility of using multimodal data to capture student engagement dynamics.

Unger, S., and Meiran, W.R. [6] developed a statistical analysis model based on survey questions meant to determine students' perceptions of online learning during the pandemic. The frequency of student responses about anxiety towards online education was examined using the chi-squared test, t-test statistical analysis, and descriptive statistics. The survey data was computed in Excel with a significance value of 0.05. The study identified goal setting and engagement as critical factors for success in online learning environments.

During the Covid-19 outbreak, Dek Ngurah Laba Laksana [7] carried out a qualitative descriptive study and data analysis to ascertain students' opinions regarding the usage of online learning in locations with spotty internet connectivity. This study covered the four stages of the research process: data collection, data reduction, data presentation, and conclusion. The categories, the standards for each category, and the connections between the categories were examined to analyze the data before offering an explanation. The study highlighted the challenges and opportunities associated with online learning in underserved areas.

T. Muthuprasad et al. [8] proposed a content analysis of student survey data to identify patterns in learners' views about online courses. A draft questionnaire was developed from a review of the literature and informal discussions with the participants in the online course. Following the collection of data regarding learners' preferences, perspectives, advantages, drawbacks, and suggestions, demographic characteristics were also collected. For most of the questions, frequency and percentage were calculated to provide an overview of the findings. The study provided insights into students' preferences, perceptions, and suggestions for improving online learning experiences.

Maphosa, V. [9] proposed performing a quantitative study to ascertain college students' perceptions of the COVID-19 e-learning deployment and the factors influencing usage. An improved version of the Unified Theory of Acceptance and Use of Technology (UTAUT) paradigm served as the foundation for the research methodology used in the quantitative study. Two internal variables and five external variables compose the model. After being chosen at random to take part in the study, 314 university undergraduate students completed the online survey. The study identified a lack of access to educational resources as a significant barrier to effective online learning.

Dewi Surani and Hamidah proposed employing content analysis to examine open-ended questions collected from qualitative interviews and quantitative data to provide a descriptive statistical analysis that shed light on students' perceptions. Based on the replies provided, a descriptive qualitative presentation of the data collected via a questionnaire on students' perspectives of online learning was made. According to the analysis, 32% of students find it difficult to understand the information, 80% of students are willing to participate in online learning, and 62% of students believe online learning offers benefits. The study revealed mixed feelings among students, with some appreciating the flexibility of online learning while others struggling with the lack of in-person interaction.

In 2022, Jamil et al. conducted a study on the potential of brain-computer interfaces (BCIs) and eye-tracking technology to enhance students' cognitive abilities in online learning. The study found that eye tracking and BCIs could be used to create neurofeedback-based interventions that would enhance learning outcomes and cognitive skills, as well as to track students' attention and cognitive engagement in real time and provide tailored feedback.

3. Design and Methodology

The Methodology of the proposed model has been depicted in Fig 1.

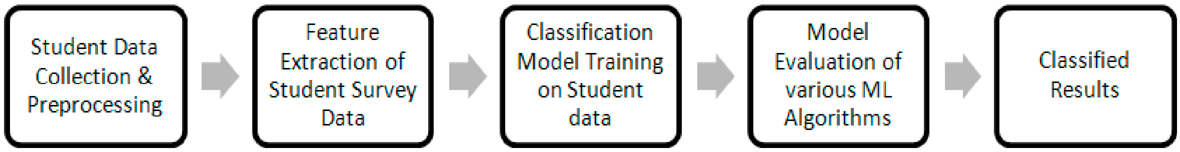

Fig.1. Architecture Framework for ML Pipeline of Student's interest towards Online Learning

3.1. Data Collection & Preprocessing

High school and engineering students were selected as the study's respondents because engineering is the most diversified discipline, with courses covering everything from social sciences to technical sciences and requiring students to work in both lab and field settings. 1006 undergraduate students from various Indian universities took part in the study. A preliminary, structured and unstructured questionnaire was developed from a review of the literature and informal discussions with the online course participants. A pre-test was conducted with ten respondents, and their feedback was considered in the development of the final questionnaire. A Microsoft Form with the final questionnaire data was shared on several student and university social media platforms. Students in Indian institutions and high schools submitted the corresponding Microsoft form using the provided questionnaire, and all of the student data was converted to an Excel form for analysis. The collected excel data was cleaned to ensure it was accurate and complete by searching for missing data, and duplicate entries, such as some students submitting the same survey form twice, were removed using the Python pandas module during the data cleaning and preparation procedures to get the data into a format that machine learning algorithms could use. The research study used the Python pandas module for data cleaning and analysis to identify and deal with missing values, outliers, and inconsistencies in the data.

3.2. Feature Extraction of Student Survey Data

Feature selection can reduce the feature dimension and enhance the model's adaptability, operational efficiency, and interpretability by choosing pertinent characteristics from among all features that are helpful to the training model. Numerous features that probe students' perceptions of online learning are included in the questionnaire created for the study, such as "Percentage of Efforts students are putting into the online learning", "Preferred Method of Teaching", "Least-Engaging Aspect of Online Learning", "Attendance Percentage", "Teachers Support", "Clarification of Doubts", "Required Resources of Online Learning" are independent features and three target features selected for evaluation are "Excitement to Attend Online Classes", "Satisfaction with Online Learning", and "Overall Rating" on a scale of 0 to 5 [1-4][7-10] and based on their association with the target variables, a subset of the most informative and relevant features was chosen. The chosen features include satisfaction ratings, perceived benefits and drawbacks, engagement levels, and learning preferences. These features could have an impact on the way students view online education. Data transformation was performed by utilizing label encoding and one-hot encoding techniques to convert categorical features into numerical representations and analyze data effectively. Features were scaled using standardization to a common range to ensure that each feature contributes equally to the analysis and avoid bias towards features with larger scales and employed methods like correlation analysis to assess the relevance and quality of the retrieved features. Principal Component Analysis was used on the dataset to reduce the number of features in order to enhance computational efficiency and reduce noise in the data.

3.3. Classification Model Training on Student Data

With 70% of the data on the training side to train the ML model, 15% of the data on the validation side to fine-tune the model's hyper parameters and 15% on the testing side to evaluate the model's performance on the unseen data, the dataset was split into training, validation and testing sets. Model training is the process of feeding data into a model and facilitating it to find patterns in the data. After the model has been trained, it can be evaluated using a held-out test set. This makes it easier to assess the way the model performs with fictitious data. Machine Learning classifiers could be used to identify the most important factors that students consider when choosing whether or not to enroll in an online course. Traditional machine learning techniques, like Random Forest [4], K-Nearest Neighbors, Support Vector Machines, and XGBoost, are chosen for the model training session, and features that are extracted from the student survey data are used as feature data to train the machine learning models for classification. Random forest [4] is a tree-based ensemble learning technique that generates a class representative of the mode of the classes to be categorized by training a large number of decision trees. Support vector machines [3] work by finding the hyperplane that divides two data classes the most effectively. SVM is well-suited to handle high-dimensional data, which is commonly observed in student perception datasets, because it incorporates numerous features that indicate student demographics, survey responses, and online learning activity. SVM is a good option for evaluating student experiences, which could include errors or missing information because of its relative insensitivity to noise and outliers in the data. This could reduce the dimensionality of the data and improve the overall performance of the ML model. KNN [3] associates a new data point with the class that is most prevalent among its K neighbors after determining which K of the training set's data points are most similar to it. XGBoost employs a gradient-boosting architecture in its tree-based ensemble learning technique. The four ML classifiers were trained on the dataset using a 10-fold cross-validation procedure.

3.4. Practical Implications

After applying machine learning algorithms, we need tools to determine how well they performed their jobs. These tools are called performance evaluation metrics. By employing several commonly-used metrics for classification problems, this work aims to undertake a comparison analysis and provide insightful information about algorithm performance. These metrics are the accuracy, precision, recall and f1-score. We have evaluated each training model using appropriate evaluation metrics, such as accuracy, precision, recall, and F1-score, concerning the unseen testing set. The most popular and possibly best metric for assessing an algorithm's effectiveness in classification challenges is accuracy. It is the proportion of correctly identified data items to all observations. Precision indicates the percentage of observations that are truly positive among those that an algorithm has projected to be positive. Recall refers to the percentage of truly positive observations that the algorithm correctly predicted. F1-score is the harmonic mean of both precision and recall. Due to the imbalanced nature of the dataset, where students expressing interest in online education significantly outnumbered those who did not, we employed weighted average metrics to evaluate the performance of our machine learning models. This approach ensured that the evaluation was not skewed towards the majority class and provided a more accurate representation of the model's ability to correctly classify students with both interest and no interest in online education. These weighted average precision, recall, and F1-score metrics provided a more balanced assessment of the models' performance across all classes. The metrics for the performance evaluation of the ML algorithms are tabulated in Fig 2.

Machine Learning Algorithms	Accuracy	Precision	Recall	F1-Score
Random Forest	93%	0.93	0.92	0.92
K-Nearest Neighbors	87%	0.88	0.87	0.873
Support Vector Machines	78%	0.79	0.78	0.78
XGBoost	92%	0.92	0.91	0.91

Fig 2. Performance Evaluation Metric Values of ML Algorithms

3.5. Major Research Findings and Classified Results

To forecast the outcome, we evaluated the output using the three target classes. The dataset was exceptional in revealing the opinions of the students regarding teaching pedagogy and online learning based on the features selected. We used grid search optimization to fine-tune each model's hyperparameters to identify the optimal combination of parameters that maximizes the model's performance on the validation set. We have evaluated a loop of different hyperparameters, including learning_rate determines the step size that allows the algorithm to learn quickly, max_depth defines the maximum depth of the trees in a random forest required to achieve a desired level of accuracy without overfitting, and n_estimators determine the number of trees in a random forest required to achieve a desired level of accuracy. A comparison between the algorithms can be used to determine the categorized result. The Random Forest classifier outperformed the other three classifiers, according to the plot's results. The random forest classifier is a well-known ensemble learning method that is robust against noise in data. The Random Forest classifier does rather well, with good recall, accuracy, and F1 scores of 0.92, 0.93, and 0.92, respectively. It appears from this that the Random Forest classifier does a good job of correctly identifying students who are interested in online learning and accurately identifying all interested students. The XGBoost classifier also does well, scoring 0.92 for precision, 0.91 for recall, and 0.91 for F1. On the other hand, compared to the XGBoost classifier, the random forest classifier consistently identifies students who are interested in online learning more accurately. Given their good performance with accuracies ranging from 0.78 to 0.87, the other two classifiers might also be a good choice in certain circumstances for predicting students' interest in online learning. The SVM classifier might be a good choice for predicting students' interest in online learning when working with high-dimensional data. The KNN classifier is a good choice for estimating students' interest in online learning when the dataset is small. Using the Python matplotlib module, a plot was made to show the model with the highest accuracy by comparing the accuracies of the various classifiers used in our model on the y-axis and their implementation on the x-axis.

Plotting the accuracy of the previously discussed classifiers is presented in Fig 3.

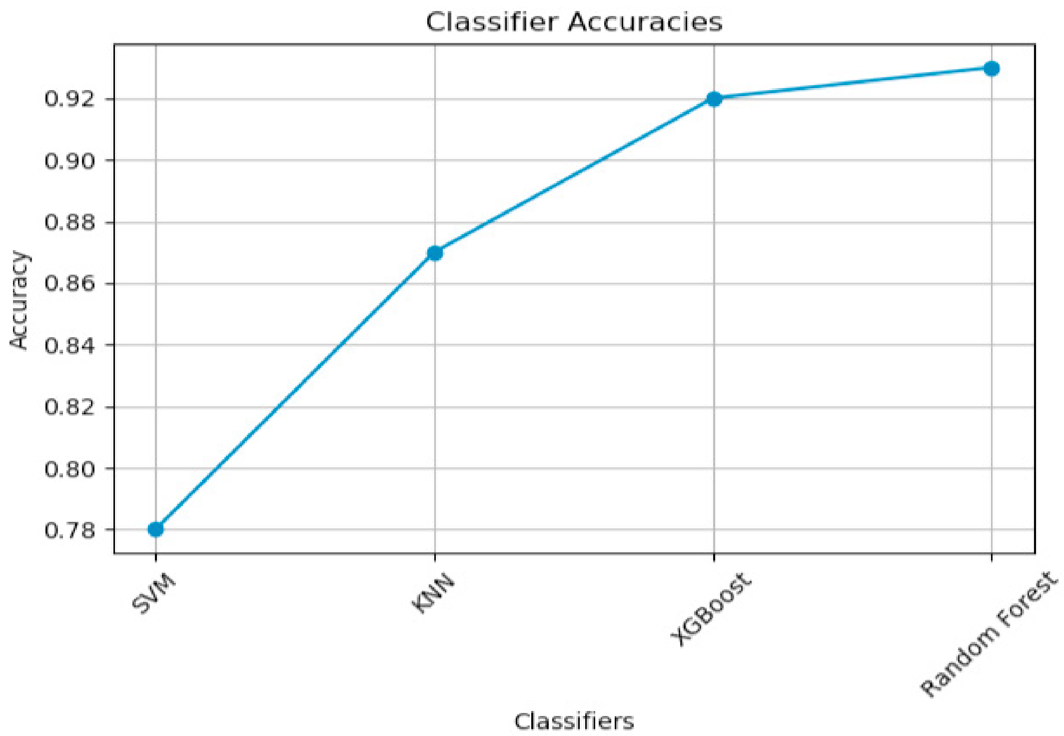

Fig 3. Accuracy Plot for ML classifiers

4. Research Limitations

A constraint of the research is that it was carried out on a comparatively small sample size of 1,006 students. The fact that the study merely included data from Indian students is another limitation. As a result, it's possible that the findings cannot be applied to students in other countries. Only self-reported survey data were used during the research.

It implies that bias may exist in the data. Future research could address the limitations of the current study by using a larger sample size, collecting data from students globally, and merging data from other sources, like learning management systems and student information systems.

5. Conclusion and Future Work

This research paper presents a dataset on student opinions and interest in virtual learning. We conducted extensive experiments to assess the effectiveness of various machine learning algorithms, including XGBoost, Random Forest, KNN, and SVM on the student dataset. As per the findings, Random Forest and XGBoost surpassed other machine learning algorithms in every metric, with Random Forest exhibiting the lowest execution time among them. Out of 1006 students, only 40% claimed they were satisfied with online education; nevertheless, they also stated that process improvement is required to improve teaching pedagogy and the quality of the learning environment. According to many students, the least engaging aspects of online learning are Project Experience (LABS), the shortage of infrastructure that is offered by traditional learning, and less exposure to practical skills. This information could be used to develop online courses that better meet the needs of students. The findings of this study give educators and e-learning platforms useful knowledge about the way students view digital learning at higher education institutions and its current state. Future studies could evaluate the students' behavior analysis and investigate their mental behavior regarding stress and social anxiety by analyzing the students' facial expressions or EEG signals in response to their online education.

References

[1] Boca, G. D. (2021). Factors influencing students' behavior and attitude towards online education during COVID-19. *Sustainability*, *13*(13), 7469.

[2] Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, *8*(3), 221-231.

[3] Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., ... & Chen, P. K. (2022). Predicting students' performance in e-learning using learning process and behaviour data. *Scientific Reports*, *12*(1), 453.

[4] Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. *Sustainability*, *14*(22), 14795.

[5] Altuwairqi, K., Jarraya, S. K., Allinjawi, A., & Hammami, M. (2021). Student behavior analysis to measure engagement levels in online learning environments. *Signal, image and video processing*, *15*(7), 1387-1395.

[6] Unger, S., & Meiran, W. R. (2020). Student attitudes towards online education during the COVID-19 viral outbreak of 2020: Distance learning in a time of social distance. *International Journal of Technology in Education and Science*, *4*(4), 256-266.

[7] Laksana, D. N. L. (2020). Implementation of online learning in the pandemic covid-19: Student perception in areas with minimum internet access. *Journal of Education Technology*, *4*(4), 502-509.

[8] Muthuprasad, T., Aiswarya, S., Aditya, K. S., & Jha, G. K. (2021). Students' perception and preference for online education in India during COVID-19 pandemic. *Social sciences & humanities open*, *3*(1), 100101.

[9] Maphosa, V. (2021). Factors influencing student's perceptions towards e-learning adoption during COVID-19 pandemic: A developing country context. *European Journal of Interactive Multimedia and Education*, *2*(2), e02109.

[10] Surani, D., & Hamidah, H. (2020). Students perceptions in online class learning during the Covid-19 pandemic. *International Journal on Advanced Science, Education, and Religion*, *3*(3), 83-95.

[11] Ismail, S. N., Hamid, S., Ahmad, M., Alaboudi, A., & Jhanjhi, N. (2021). Exploring students' engagement towards the learning management system (LMS) using learning analytics. *Computer Systems Science & Engineering*, *37*(1).

[12] Jamil, N., Belkacem, A. N., & Lakas, A. (2023). On enhancing students' cognitive abilities in online learning using brain activity and eye movements. *Education and Information Technologies*, *28*(4), 4363-4397.

[13] Zhang, G., Zhu, Z., Zhu, S., Liang, R., & Sun, G. (2022). Towards a better understanding of the role of visualization in online learning: A review. *Visual Informatics*.

[14] Boling, E. C., Hough, M., Krinsky, H., Saleem, H., & Stevens, M. (2012). Cutting the distance in distance education: Perspectives on what promotes positive, online learning experiences. *The Internet and Higher Education*, *15*(2), 118-126.

[15] Szopiński, T., & Bachnik, K. (2022). Student evaluation of online learning during the COVID-19 pandemic. *Technological Forecasting and Social Change*, *174*, 121203.