PREDICTING STUDENTS' ADAPTABILITY IN ONLINE EDUCATION THROUGH ENSEMBLETECHNIQUES

Abstract

Learning is a lifetime process for individuals, and learning takes place from infants to the elderly. Learners gather information, recognise it, map it to knowledge, and apply it in their daily lives.Since the Pandemic was released into the world; online education has become a buzzword. To continue their educational activities, the majority of educational institutions went online. An online eduction is nearly complete when the teaching and learning community has access to the necessary digital infrastructure, including smart phones, tablet computers, and the most up-to-date software platform.Students' ability to adjust to online education varies depending on their gender, device, financial situation, and other factors. As a result, educational institutions face challenges in identifying students' online aptitude levels and delivering exceptional instruction via ICT. In this study, we gathered evidence on 1205 students, including their age, gender, educational level, device used for online education, financial condition, internet type, network type, and so on. To acquire a sense of the effectiveness of online education, we used different machine learning methods, including Decision Tree (DT) and ensemble bagging approaches, on our dataset to predict the students' level of adaptation to online learning. The Ensemble bagged trees approach has the highest accuracy of 89.8 percent and outperformed the other algorithm.

Keywords: Online Education, Ensemble, Machine Learning, Bagging ,AUC

Author

S. Chithra
Assistant Professor
Christ Academy Institutes for Advanced Studies
Bengaluru, Karnataka, India.

I. INTRODUCTION

The advent of technology allows us to create online education systems, which provide a solution to virtual or remote learning issues. Given the current scenario, numerous aspects of education are shifting to digital platforms. In light of these changes, students are faced with the challenge of accepting online education. In the following discussion, we present an overview of the findings from a review of important studies on online education.
The researcher's study highlights the global concern posed by COVID-19 on education systems. With over 100 countries closing schools due to the pandemic, the study reveals severe challenges in maintaining effective online education. The research identifies a range of barriers hindering student-teacher interaction for continued learning during lockdowns. These barriers include lack of digital skills and infrastructure in rural areas, technological limitations, personal and domestic challenges, institutional obstacles, communication difficulties, inadequate electricity supply, network issues, insufficient training, financial constraints, and resistance to change. The study aims to understand how students are adapting to online education during the pandemic[1] and [2].

II. LITERATURE SURVEY

The researchers conducted a study aimed at enhancing the online education model. In a study by Rojan et al. [2], significant improvements were observed in students' performance and satisfaction, underscoring the advantages of online education. The study revealed comparable performance levels and student satisfaction between off-campus and on-campus settings. Another study by William et al. [4] concentrated on formative assessment to enhance learning. The preliminary findings indicated that 85% of students reported increased learning through online education. The primary focus of the researchers was to enhance assessment methods for both students and teachers, including self-assessment and peer assessment.

Rolim et al. [5] utilized a Supervised Machine Learning algorithm to detect positive practices by analyzing feedback from Learning Management System (LMS) courses. Simultaneously, William et al. [4] focused on improving the Online Education Model by integrating Machine Learning and Data Analysis into a Learning Management System (LMS).

Researchers compared five categorization methods for predicting student achievement in study [6]. For nominal data, they used three variants of the nave bayes technique, while for numerical data, they used multiple regression and support vector machine algorithms. This study was notable for its use of the naive Bayes method.

In a separate study [7], the authors analyzed student performance prediction in final exams, specifically in the context of distance education. They assessed the efficacy of six distinct machine learning algorithms: decision trees, neural networks, naive bayes, instance-based learning, logistic regression, and support vector machines. These algorithms were then juxtaposed against a genetic algorithm rooted in the induction of decision trees.

Similarly, researchers in [8] compared various strategies for measuring student understanding through intelligent tutoring systems. At the action level, they also used ensemble approaches to combine numerous student models. Evaluations were carried out to measure future performance projections, both within the tutoring system and through paper-based post-tests.

Ensemble learning techniques have received increased attention in the field of predictive modelling. By combining several learning algorithms, these strategies prove to be quite effective in improving overall prediction accuracy[9].

In reference [10], authors introduced a comprehensive Student Success System (S3) encompassing an ensemble-based analytical framework designed to monitor students' academic achievements. To identify at-risk children, the system includes a versatile predictive modelling engine that use machine learning methodologies. It also has advanced data visualisations and a case management tool for putting intervention methods into action.

In a separate study, Mingjie et al. [11] delved into the concerning issue of dropout rates in online education and E-learning courses. Their focus centered on predicting viable strategies to prevent student attrition.

In the context of online education, Xiaofeng and colleagues [12] explored particular student variables for estimating pass rates. Their study sought to predict student success rates and to discover the most effective machine learning algorithm for detecting critical student attributes that influence learning outcomes. To create a feature model, they used decision trees (DT), support vector machines (SVM), and deep neural networks (DNN).

III. METHODOLOGY

1. **Machine Learning:** Machine learning involves training machines to respond to specific inputs or situations by leveraging prior learned inputs. The primary objective of machine learning is to comprehend underlying patterns and integrate them into models accessible and interpretable by humans. In contrast, machine learning techniques enable computers to learn from data inputs and employ mathematical analysis to generate values within defined ranges. Through machine learning, computers can create models from sample data to automate decision-making based on information inputs. This capability enables computers to act autonomously, eliminating the need for explicit instructions thanks to machine learning.

   Numerous machine learning algorithms are available to forecast potential outcomes of student adaptability levels. Our dataset underwent training and testing using a range of these algorithms. Specifically, we employed K-Nearest Neighbor, Decision Tree, and Ensemble techniques for prediction and analysis purposes.

2. **Decision Tree:** The Decision Tree Classifier was employed as a fundamental predictive model in our study. A Decision Tree (DT) is constructed with a flow-chart tree structure, where internal nodes signify feature tests, and leaf nodes correspond to the ultimate output. This algorithm is widely adopted due to its swift and efficient approach to categorizing datasets, offering simplicity in interpretation and implementation compared to alternative classification methods.The architecture of the DT begins with a root node and proceeds to segment data into progressively smaller subsets containing instances with

akin attributes. Within the DT algorithm, entropy frequently serves to gauge the significance of homogeneous student characteristics. The ensemble of decision trees adheres to the principle of recursive partitioning, with each tree structured based on iterative data segmentation[13].

3. **K-Nearest Neighbourhood:** Out of the numerous machine learning algorithms available, K-Nearest Neighbors (KNN) stands out as one of the most straightforward methods. KNN's popularity can be attributed to its ease of interpretation and the relatively low computational time required for its calculations [14]. In the KNN approach, the algorithm first determines the value of 'k,' which represents the number of neighboring data points to consider. Subsequently, it computes the distance function for the selected 'k' neighbors and uses this information to assign the class label that appears most frequently among these k nearest neighbors.

4. **Ensemble Bagging:** Bagging [15] is based on the Bootstrap [16] sampling approach. Each iteration generates a new set of bootstrap samples to create independent classifiers of the same algorithm. Bootstrap sampling includes randomly choosing data points with replacement, implying that some occurrences might be repeated while others may be excluded from the original dataset during sampling.

   The classifiers built in the previous stage are combined in the next phase of bagging. This combination is then put through a voting system to provide a final forecast. Bagging, as described in [15], emerges as a potent ensemble technique, particularly suited to volatile learning algorithms. Due to modest modifications in the training dataset, such algorithms demonstrate substantial shifts in predictions. Decision Trees and Neural Networks are two examples.

5. **Dataset Description:**

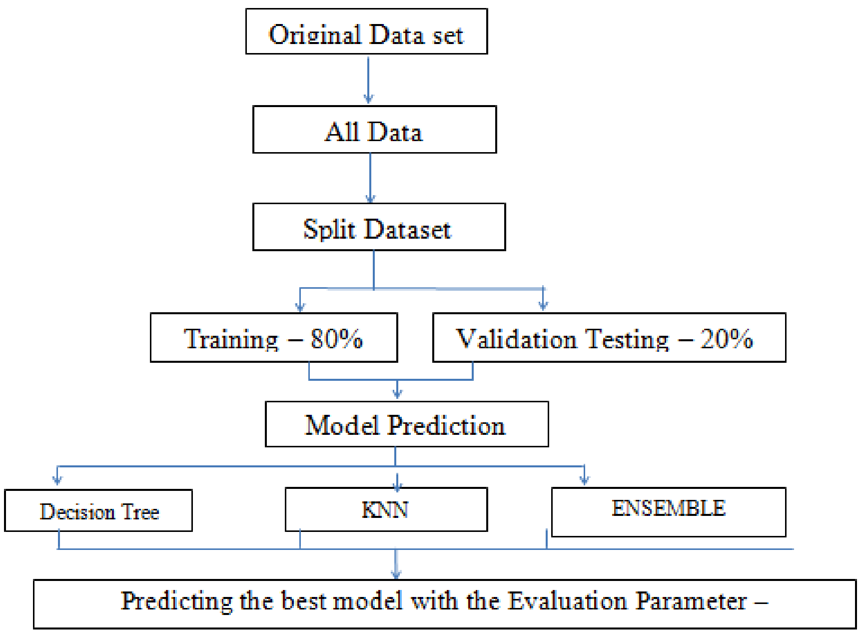

Figure 1: Flow model of the proposed methodology.

The dataset for this study was obtained using a combination of online and offline surveys. We acquired student data from numerous educational levels, including universities, high schools, and colleges. Our study collected a total of 1205 records between December 10, 2020 and February 5, 2021. This dataset includes 14 distinct attributes, including age, gender, educational level, institution type (government or non-government), geographical location, IT student status, educational background, level of exposure to power outages, internet quality, class schedules, the family's financial situation, the type of device used to attend classes, and the availability of the institution's dedicated Learning Management System (LMS).The flow model for the suggested methodology is shown in Fig1

6. **Model Prediction:** The proposed study incorporates three classification algorithms—Decision Tree, KNN and Ensemble bagging techniques applied to the datasets. The comparison of outcomes is based on evaluation metrics like Accuracy, AUC. After implementing the algorithmic approach, performance assessment involves a diverse range of metrics, facilitating comparative analysis. Model accuracy serves as a crucial determinant of its proficiency in accurately categorizing data.

   Specifically the AUC-ROC curve, provides a visual representation of the performance of our machine learning classifier. While primarily designed for binary classification tasks, we will explore how this concept can be extended to assess multi-class classification problems.

   The interpretation of AUC values is as follows:

   AUC Value >= 0.9: The model is considered excellent.
   AUC Value between 0.8 and 0.9: The model is deemed good.
   AUC Value between 0.7 and 0.8: The model is classified as fair.
   AUC Value between 0.6 and 0.7: The model's performance is seen as poor.
   AUC Value < 0.6: The model is regarded as very poor.

IV. RESULTS AND DISCUSSION

The research study is broken down into many stages to determine how well students can adjust to online learning and how to evaluate the findings. However, evaluation findings concerning the model's performance accuracy were taken into account. Comparing the performance of the classifiers was the main goal of the study. The algorithm with the best predictive model will be employed to carry out the prediction of the cirrhosis disease. Table 1 displays the performance results for each classifier that was considered, as well as their visual performance.

Table 1: Performance Evaluation

ML Algorithms	Accuracy	AUC
Tree	85.70%	0.97
KNN	49.80%	0.64
Ensemble Bagging	89.80%	0.99

A comprehensive depiction of the experiment's classifiers' performance accuracy is presented in intricate detail within the subsequent figures.

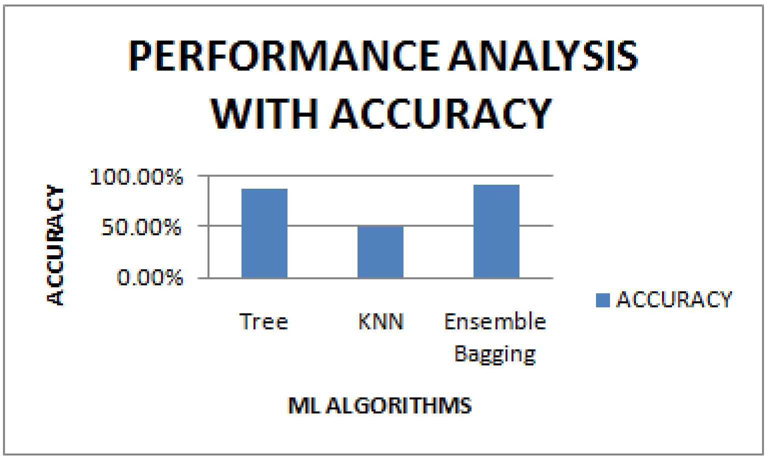

Figure 2: Comparison of machine algorithms according to accuracy

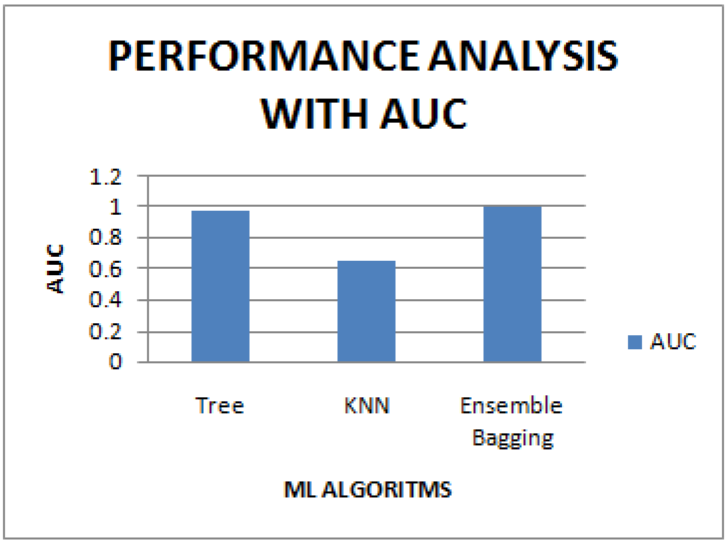

Figure 3: Comparison of machine algorithms according to AUC

V. CONCLUSION

In this article, we used Machine Learning approaches to forecast students' level of adaption to online education. Our data collection included both online and offline surveys administered at various educational levels. K-Nearest Neighbour, Decision Tree, and Ensemble Bagging are among the approaches used. The Machine Learning models performed well in terms of prediction.Among the models used, the Ensemble Bagging Classifier had the highest prediction accuracy, at 89.80%. These algorithms produced a variety of results depending on accuracy and AUC evaluations. These strategies' efficacy was extensively compared and analysed.This study's consequences extend to education sector decision-makers, providing insights into the current condition of online education and the extent to which students have acclimated to it.

REFERENCES

[1] E. M. Onyema, N. C. Eucheria, F. A. Obafemi, S. Sen, F. G. Atonye, A. Sharma, and A. O. Alsayed, “Impact of coronavirus pandemic on education,” Journal of Education and Practice, vol. 11, no. 13, pp. 108– 121, 2020.

[2] R. E. Baticulon, J. J. Sy, N. R. I. Alberto, M. B. C. Baron, R. E. C. Mabulay, L. G. T. Rizada, C. J. S. Tiu, C. A. Clarion, and J. C. B. Reyes, “Barriers to online learning in the time of covid-19: A national survey of medical students in the philippines,” Medical science educator, pp. 1–12, 2021.

[3] R. Afrouz and B. R. Crisp, “Online education in social work, effectiveness, benefits, and challenges: A scoping review,” Australian Social Work, vol. 74, no. 1, pp. 55–67, 2021.

[4] D. Wiliam, “Assessment in education: Principles, policy & practice,” Assessment in Education: Principles, Policy and Practice, vol. 15, no. 3, pp. 253–257, 2008.

[5] A. P. Cavalcanti, R. F. L. de Mello, V. Rolim, M. Andre, F. Freitas, ´ and D. Gasevic, “An analysis of the use of good feedback practices in ˇ online learning courses,” in 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), vol. 2161. IEEE, 2019, pp. 153–157.

[6] Wilhelmiina, H. and Vinni, M. 2006. Comparison of machine learning methods for intelligent tutoring systems,Intelligent Tutoring Systems, Vol. 4053, of Lecture Notes in Computer Science, pages 525-534.

[7] Kalles, D. and Pierrakeas C. 2004. Analyzing student performance in distance learning with genetic algorithms and decision trees, Laboratory of Educational Material and Educational Methodology Hellenic Open University, Patras, Greece.

[8] Weka, University of Waikato, New Zealand,http://www.cs.waikato.ac.nz/ml/weka/

[9] Dietterich TG.(2000). Ensemble methods in machine learning. In: Proceedings of Multiple Classifier System. vol. 1857. Springer;(2000). pp. 1–15.

[10] M., Bahador B. Nooraei., and Heffernan, Neil T.2011. Ensembling Predictions of Student Knowledge within Intelligent Tutoring Systems, Joseph A. Konstan et al. (Eds.): UMAP 2011, LNCS 6787, pp. 13–24, 2011.© Springer-Verlag Berlin Heidelberg .

[11] M. Tan and P. Shao, “Prediction of student dropout in e-learning program through the use of machine learning method.” International Journal of Emerging Technologies in Learning, vol. 10, no. 1, 2015.

[12] X. Ma, Y. Yang, and Z. Zhou, “Using machine learning algorithm to predict student pass rates in online education,” in Proceedings of the 3rd International Conference on Multimedia Systems and Signal Processing, 2018, pp. 156–161.

[13] “Machine learning 101: Decision tree algorithm for classification,” Available Online: https://www.analyticsvidhya.com/blog/2021/02/machinelearning-101-decision-tree-algorithm-for-classification/, Access Date: 10/05/2021.

[14] “Understanding math behind knn (with codes in python),” Available Online: http://www.datascribble.com/blog/machinelearning/understanding-mat

[15] Breiman L.1996. Bagging predictors, Machine Learning, 24(2):123-140.

[16] Efron, B. and Tibshirani, R.J. 1993. An Introduction to the Bootstrap, Chapman & Hall, New York.