Educational Data Mining (Students' Flexibility in Online Education)

Bhagyalakshmi Katta

ABSTRACT

The COVID-19 pandemic has led to a shift towards online education as the primary mode of learning. As a result, it is crucial to assess the effectiveness of this form of instruction and students' learning flexibility. In this study, educational data mining approaches were employed to investigate the flexibility of students in online learning. The dataset includes various variables such as age, gender, education level, IT student, institution type, location, load-shedding, and others. To predict the adaptability level of online students using this data, Random Forest, Decision tree, and Support Vector Machine models were used. The study involved evaluating the F1 score, accuracy, recall, and precision to validate the effectiveness of the designed model. The Random Forest algorithm demonstrated a high accuracy of 91% in predicting the adaptability level compared to the Decision Tree and Support Vector Machine algorithms. Thus, the Random Forest algorithm may be a valuable tool for evaluating the adaptability level of online students. The results of this study could raise the standard for online learning and make sure that students are equal to the challenges posed by this type of instruction.

INTRODUCTION

The global outbreak of COVID-19 has had an extensive and far-reaching effect on education, causing numerous educational establishments to transition to virtual learning as their main form of teaching. Although online education has been utilized for some time, it is now the only feasible solution for a lot of pupils due to the pandemic. Despite this, electronic instruction brings about a handful of difficulties both for learners and instructors alike, such as a shortage of face-to-face communication, technical problems and necessitating stronger self-management abilities and better time management skills. It is vital to assess the effectiveness of online learning and identify factors that support students' adaptability and versatility in this learning environment to overcome these challenges. Large datasets can be examined to find patterns and connections that can improve educational outcomes using educational data mining. The scope of the study is restricted to data analysis and does not include data collection. To stop the COVID-19 epidemic from spreading, many educational institutions had to transition to online instruction. The educational environment has undergone major changes because of this transformation in education delivery strategies.

The research question for this study is “What factors affect students' adaptability in online learning, and how can their level of adaptability be predicted using educational data mining techniques?” This research objective is to analyze the elements that shape students' ability to adjust in online education and create predictive models with Random Forest, Support Vector Machine and Decision tree algorithms to measure learners' adaptability in this type of learning. The significance of this study lies in its potential to aid in the creation of successful online learning environments that support student success and flexibility. In this study, the analysis of data from a single educational institution in a specific area is the focus.

The outline of the paper:

Abstract: It gives an overview of work including research background, methodology, findings, and results.
Introduction: This section offers background information, a problem statement, and a list of any literature gaps. It also has significance, contribution, scope, and limitations.
Methodology: It explains the data analysis process, including the dataset and analytical methods.
Results and Discussions: It presents the findings of the data analysis, as well as the elements that affect the adaptability of learners in online learning.
Conclusion: It offers a conclusion and suggestions for additional investigation.

METHODOLOGY

The dataset used for this project has to do with the adaptability of online students. The dataset includes a few features, including age, gender, institution type, education level, location, IT student, internet type, device, load shedding, network type, class duration, self lms, financial condition, and adaptivity level. There are 1205 records in the collection overall, and there are 14 features.

Data Collection:

For research, the pre-processed dataset from the Kaggle is taken. 1205 data instances and 14 attributes were merged to predict the adaptability level of online students. Figure 1 is a pie chart showing the adaptability level of online students that is low, moderate, and high.

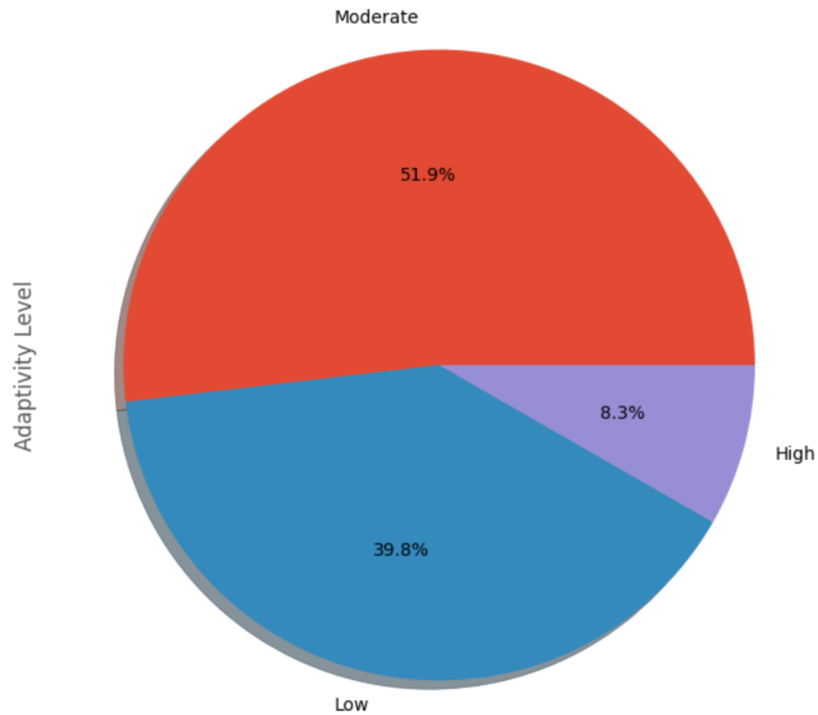

Figure 1: pie chart showing the adaptability level of online students.

Data Pre-Processing:

Pre-processing of data typically entails feature engineering, data cleansing, and formatting. There were no duplicates or missing data in the 1205 online students records that make up the pre-processed dataset used. The results show that 100 records have a 0 value, which indicates high level, 450 records have a 1 value, which indicates low level, whereas 625 records have a 2 value, which indicates moderate level. The dataset was also split between training and testing sets using a 75:25 split.

Data Analysis:

The characteristics that were shown to be most crucial for predicting adaptability level of online students were chosen using a correlation matrix. The most important variables were found to be gender, age, education level, network type, institution type, internet type, class duration, self-lms, financial situation, device, and adaptivity level. Random Forest has been selected as the ML method for predicting adaptability level of online students. This technique was chosen because it is frequently employed for binary classification tasks and because its implementation is quite straightforward. The system was tested using measures including accuracy, precision, and recall after being trained on the training set.

Model Evaluation:

Accuracy, precision, and recall were the evaluation metrics employed in the study.

RESULTS AND DISCUSSIONS

Machine learning techniques like Decision Tree, Support Vector Machine, and Random Forest are frequently used for classification problems. These algorithms are helpful in predicting pupils' levels of adaptability and determining whether they are high or low. The quality and size of the dataset, as well as the optimization of the algorithm's hyperparameters, all affect how accurately these algorithms estimate the adaptability level. To examine the effectiveness of these machine learning algorithms in predicting the level of adaptability of online students using educational data mining, the dataset in this study was randomly divided into training (75%) and testing (25%) sets. Overall, all three machine learning models performed optimally when used to forecast how adaptable online students will be. With a precision score of 90.0% and a recall score of 86.0%, the Random Forest model earned the greatest accuracy score of 91.0% in predicting the adaptability level. The accuracy, precision, and recall scores for the decision tree model were 90.0%, 88.0%, and 85.0% respectively. The Support Vector Machine model scored 71.0% for accuracy, 74.0% for precision, and 67.0% for recall. In conclusion, utilizing educational data mining, all three machine learning models are effective predictors of the adaptability level of online students. The Decision Tree method obtained a little higher precision score than the others, but the Random Forest model outperformed the others slightly in terms of accuracy and recall. All three evaluation measures yielded very poor results for the Support Vector Machine model.

CONCLUSION

The aim of this study was to compare the accuracy of three ML algorithms, namely Random Forest, Support Vector Machine, and Decision tree, in predicting the factors influencing the adaptability level of online students. The dataset used for this purpose was downloaded from Kaggle, and prediction models were created using the three algorithms. The performance of these models was then compared using evaluation metrics like F1 score, accuracy, recall, and precision. The results of this investigation showed that the Random Forest algorithm outperformed the Decision Tree and Support Vector Machine methods in every measure, with an accuracy score of 91%, followed by Decision Tree with 90%, and Support Vector Machine with 71%.

In conclusion, the Random Forest algorithm is recommended for forecasting the adaptability level of online students, over Decision Tree and Support Vector Machine methods. It would be interesting to explore the performance of other ML algorithms such as logistic regression and neural networks in predicting the adaptability level of online students for educational data mining. The findings of this study add to the conversation around online learning and provide insights for educational institutions to improve students' responsiveness to online learning.

REFERENCES

1. Kotsiantis, S. B. (2005). Applications of machine learning in real-life problems. International Journal of Emerging Technologies in Learning, 1(1), 20-23.
2. Wang, Y., Chen, N. S., & Liang, T. H. (2011). The effects of aligning learning tasks with learner's aptitude on course performance in an online learning environment. Computers & Education, 57(3), 2214-2226.
3. Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1), 368-384.
4. Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3-17.
5. Ochoa, X., & Duval, E. (2009). Quantitative evaluation of the use of learning objects. Journal of Educational Technology & Society, 12(4), 140-152.
6. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE review, 46(5), 30-40.
7. Chen, X., & Wu, H. (2015). The effects of online learning on students' course outcomes: A meta-analysis. Computers & Education, 78, 27-38.
8. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.
9. Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. Computers & Education, 68, 458-472.
10. Wu, J. H., Tennyson, R. D., & Hsia, T. L. (2010). A study of student satisfaction in a blended e-learning system environment. Computers & Education, 55(1), 155-164.