

Analysis and Prediction of Students' Adaptation to Online Education Systems Based on Data Analysis and Decision Tree Machine Learning Algorithms

Yucong Li

College of Engineering, Northeastern University, Boston, Massachusetts 02115, USA

li.yuco@northeastern.edu

Abstract. In today's digital age, the popularity and development of online education systems provide students with more flexible and convenient ways of learning. However, students' adaptation to the online education system is affected by a variety of factors, including gender, age, educational background, and field of specialisation. Through in-depth analyses and studies of these factors, the following conclusions can be drawn: gender has little influence on students' adaptation to online education, and male and female students perform similarly overall, but the proportion of male students at high adaptation levels is significantly higher than that of females. The majority of students show medium adaptability, indicating that the overall effect of online education is average. students in the age groups of 6-10, 16-20 and 26-30 years old have lower adaptability levels, and there are more low adaptability groups among students in colleges and universities. students majoring in IT are more adapted to the online education system, and students not majoring in IT have relatively poorer adaptability level. Local students are more adaptable to online education than foreign students. In areas with unstable electricity, students' adaptability is usually lower. The decision tree algorithm predictions showed good overall model accuracy, with higher prediction accuracy for students with high, low and medium levels of adaptability. The test set accuracy was 93.27%, and the precision and recall were both 93.33%, indicating excellent model predictions. In summary, by deeply analysing the influence of various factors on students' adaptation degree to online education and using the random forest algorithm to make predictions, it can provide an important reference for improving the effectiveness of online education systems and provide useful insights for personalised education.

Keywords: online education, machine learning algorithms, decision tree

1. Introduction

In today's digital era, the popularity and development of online education systems provide students with more flexible and convenient ways of learning [1]. However, the degree of students' adaptation to the online education system is affected by a variety of factors, including personal characteristics, technological capabilities, and learning styles [2]. Therefore, it is of great significance to analyse and study the various factors of students' adaptation to the online education system.

By analysing and studying the various factors of students' adaptation to online education systems, it can help educational institutions and platforms to better understand the needs and characteristics of students, so as to optimise the design and services of online education systems [3]. For example, understanding the adaptation of students of different age groups, learning backgrounds, and subject preferences to the online education system can help personalise the recommended course content or take targeted tutoring measures. In-depth analyses of the factors that contribute to students' adaptation to online education systems can also help improve teaching effectiveness and learning outcomes [4,5].

Data analytics and machine learning algorithms play an important role in studying students' adaptation to online education systems. Data analytics can help to collect and process a large amount of data on student behaviour, feedback and performance, from which potential patterns and trends can be mined [6]. Through data analysis, it is possible to build an understanding of the complex relationships between the factors that influence students' adaptation to the online education system.

In turn, machine learning algorithms can further improve the understanding of the correlations between these factors and construct predictive models. Through supervised machine learning algorithms such as decision trees and support vector machines, models can be trained based on historical data to predict the impact of different factors on the degree of adaptation of different

types of students to the online education system [7]. Analysing and studying the various factors of students' adaptation to the online education system is of great significance for improving the quality of online education and promoting personalized education. This paper analyses and researches various factors affecting students' adaptation to the online education system based on data analysis and machine learning algorithms to provide strong support for promoting the development of online education.

2. Data Set Sources and Data Analysis

2.1. Source of Data Sets

The data used in this paper comes from the Kaggle public dataset, which is available at (<https://www.kaggle.com/code/vishnu0399/adaptability-analysis-of-online-education-system>). The dataset includes information about students in various dimensions, including students' gender, age, education level, type of institution, whether they are IT majors, location, financial status, type of Internet, type of network, duration of class, and device, and each student's data corresponds to one piece of information, and each piece of information finally corresponds to the conclusion given by the student about the degree of adaptability to the online education system, which is categorised into low, moderate and high.

2.2. Data Analysis

Starting from the dimensions of students' information, the influence of each variable on students' level of adaptation to online education is analysed separately, in terms of gender, as shown in Fig. 1; age, as shown in Fig. 2; education level, as shown in Fig. 3; whether it is an IT major, as shown in Fig. 4; whether it is a local student, as shown in Fig. 5; and the local network situation, as shown in Fig. 6.

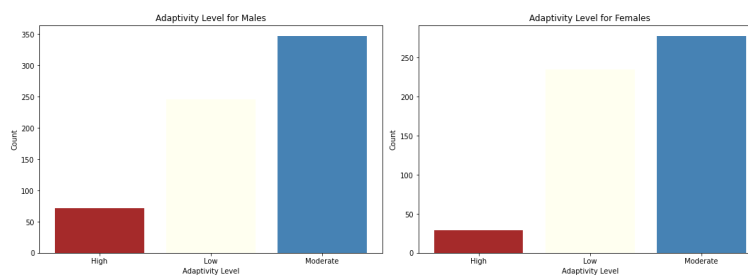


Figure 1. Classification of Statistical Results (Photo Credit: Original)

In terms of gender, the degree of students' adaptation to online education is similar for both men and women, with the majority moderately adapted, a medium number not adapted and a few very adapted. In terms of the difference in the distribution of males and females, male students show a more adaptable mode of online education than female students, and the proportion also shows that a significantly larger proportion of male students are very well adapted to online education than females.

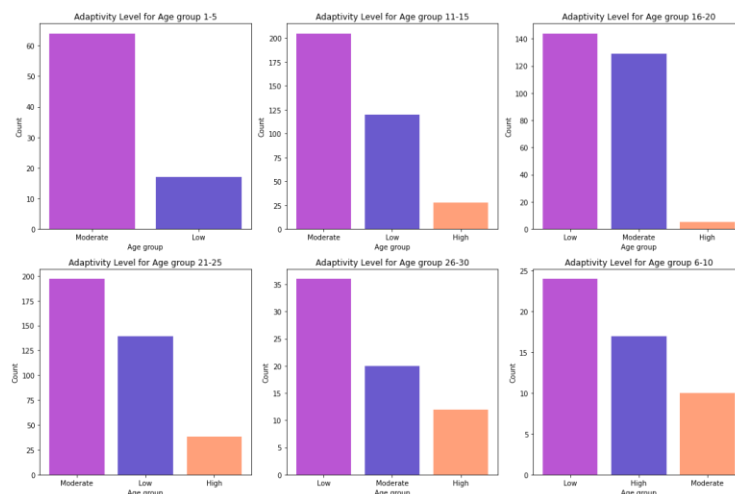


Figure 2. Classification of Statistical Results (Photo Credit: Original)

The majority of the students surveyed had a medium level of adaptation, which suggests that online education is moderately effective. the 6-10, 16-20, and 26-30 age groups had a lower level of adaptation.

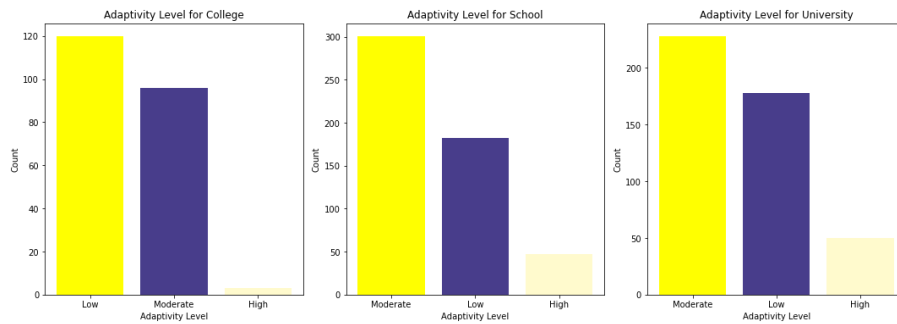


Figure 3. Classification of Statistical Results (Photo Credit: Original)

As can be seen from the above graph, in terms of educational attainment, there are more people with low adaptability in colleges and universities.

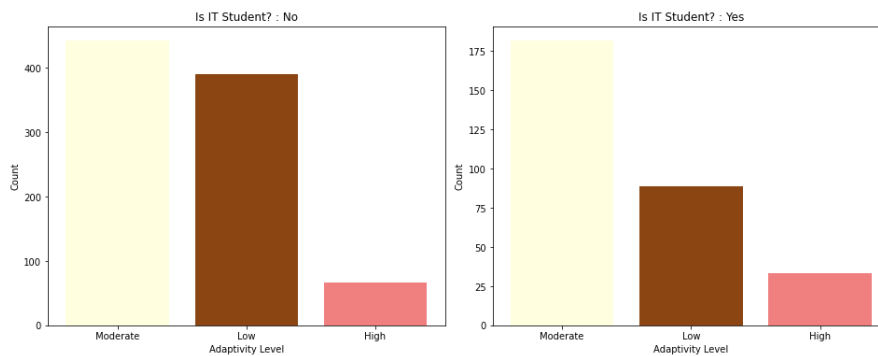


Figure 4. Classification of Statistical Results (Photo credit: Original)

Categorised by whether or not they are IT majors, observing the difference in the degree of adaptation to online education shows that IT majors are more adapted to online education, while non-IT majors are less adapted to online education.

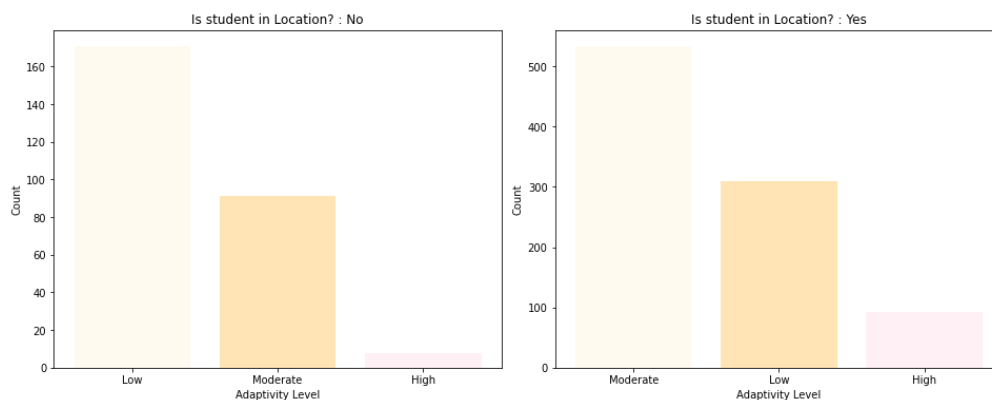


Figure 5. Classification of statistical results (Photo credit: Original)

Local students are more comfortable with online education than out-of-town students.

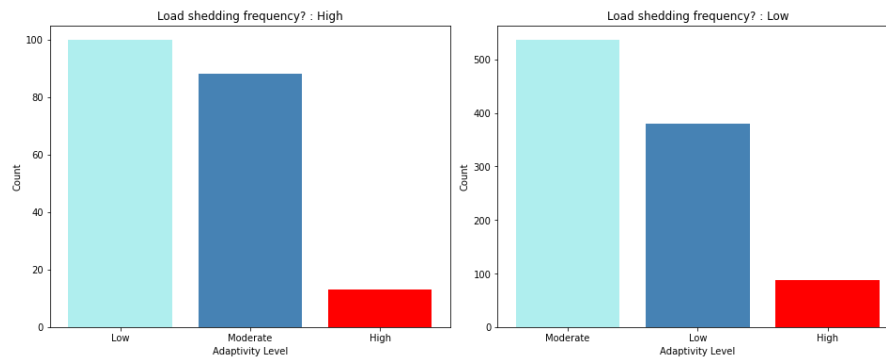


Figure 6. Classification of Statistical Results (Photo credit: Original)

Where there are frequent power outages or cuts, students are often less able to adapt due to lack of electricity and internet.

3. Decision Tree Forecasting

After exploring the influence of various factors on students' level of adaptation to online education, this paper uses the random forest algorithm of decision trees to predict students' level of adaptation. Decision tree is a commonly used machine learning algorithm that is widely used in classification and regression problems. Its principle is based on dividing the dataset recursively and predicting the values of the target variables through a series of rules [8]. The core idea of the decision tree algorithm is to represent different decision paths by constructing a tree structure, where each node represents a feature attribute, each edge represents an attribute value, and the leaf nodes represent the final classification or regression results.

The decision tree construction process consists of three key steps: feature selection, tree generation and pruning. In the feature selection stage, the algorithm evaluates the purity or uncertainty reduction of each feature after dividing the dataset, and selects the best feature as the basis for dividing the current node [9]. Commonly used feature selection metrics include information gain (ID3 algorithm), information gain ratio (C4.5 algorithm), Gini index (CART algorithm), and so on. In the tree generation phase, the algorithm recursively divides the dataset into subsets and generates the corresponding nodes until the stopping conditions are met (e.g., the maximum depth is reached, the number of node samples is less than a threshold, etc.) [10]. Finally, in the pruning stage, some methods (e.g., pre pruning, post pruning) can be used to avoid overfitting and improve the model generalisation ability. Decision trees are widely used in practical applications because of their strong interpretability, ability to handle discrete and continuous data, and insensitivity to outliers.

The training set and test set are divided according to the ratio of 6:4, and experiments are carried out using a local standalone workstation with Python 3.8, 32G of RAM, and 3080 graphics card type. The parameters of precision, accuracy, and recall are used to evaluate the prediction effect of the model and to output the confusion matrix of the model's prediction, and the results are shown in Fig. 7.

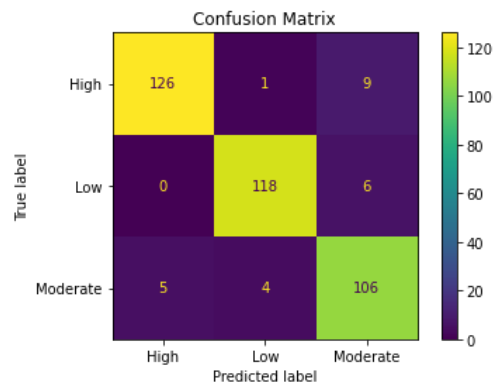


Figure 7. Confusion Matrix of the Model's Prediction (Photo Credit: Original)

From the confusion matrix, 126 students with a high degree of adaptation predicted correctly, 118 students with a low degree of adaptation predicted correctly, 106 students with a moderate degree of adaptation predicted correctly, and only 24 students predicted incorrectly, and the model's prediction accuracy was generally good. The accuracy of the test set is 93.27%, precision is 93.33%, recall is 93.33%, and the model predicts well.

4. Conclusion

By analysing and studying the multiple factors involved in students' adaptation to the online education system, the following conclusions can be drawn: gender does not have a significant effect on students' level of adaptation to online education, and both male and female students perform similarly in terms of their level of adaptation, with the majority being at a medium level of adaptation and a few showing a very high or very low level of adaptation. However, in terms of gender distribution, male students are more inclined to adapt to the online education mode than female students, especially at the high adaptation level, where the proportion of male students is significantly higher than that of female students.

The survey shows that most of the students surveyed have a medium level of adaptability, which implies that the overall effect of online education is still average. Students in the age groups of 6-10, 16-20 and 26-30 showed lower levels of adaptation. In addition, in terms of educational attainment, the low adaptation group is more prominent among college and university students.

Further observation reveals that there is a significant difference between IT majors and non-IT majors, with IT majors more adapted to the online education system, while non-IT majors show a lower level of adaptation. Local students demonstrated greater adaptation to online education compared to foreign students. It is worth noting that in areas with frequent power outages or blackouts, students usually exhibit lower adaptability due to unstable power and network.

Finally, the prediction of students' level of adaptation through the decision tree algorithm showed the overall effectiveness of the model. The confusion matrix showed that 126 students with high level of adaptation were correctly predicted, 118 students with low level of adaptation were correctly predicted, 106 students with medium level of adaptation were correctly predicted, and only 24 were incorrectly predicted in the test set. The accuracy of the model reached 93.27%, and the precision and recall were both 93.33%, which proved that the model effectively predicted students' online education adaptation level and achieved good results.

References

- [1] Segbenya, M., Minadzi, M. V., Bervell, B., et al. (2024). Online teaching intention among distance education course tutors: Modeling the effects of human resource factors and moderating role of gender. *Computers in Human Behavior Reports*, 13100380-.
- [2] Ortagus, C. J., Hughes, R., Allchin, H. (2024). The role and influence of exclusively online degree programs in higher education. *American Educational Research Journal*, 61(2), 404-434.
- [3] Top platform expands its reach: Now offering online SEO education globally. M2 Presswire, 2024.
- [4] Guo, C., Xu, Z., Fang, C., et al. (2024). China survey report on high schools' online learning status during the pandemic. *ECNU Review of Education*, 7(1), 182-194.
- [5] Top platform expands its reach: Now offering online SEO education globally. M2 Presswire, 2024.
- [6] Tian, H., Sun, M., Yin, Z., et al. (2024). Developing an evaluation index system for the online learning literacy of physical education teachers in China. *Frontiers in Psychology*, 15.
- [7] NC State College of Education ranked #27 in U.S. News' Best Online Programs Rankings. M2 Presswire, 2024.
- [8] Rajaraman, G., Klein, R., Sinnayah, P. (2024). Zoomed in, zoned out: Academic self-reports on the challenges and benefits of online teaching in higher education. *Education Sciences*, 14(2).
- [9] McLean, L., Bullivant, C., Moeke, T., et al. (2024). User-led learning preferences to inform rapid learning online education supporting evidence-based best practice in oncology. *Studies in Health Technology and Informatics*, 310, 1530-1531.
- [10] Forsyth, R., Amon, L. K., Ridout, B., et al. (2024). Health professionals' use of online communities for interprofessional peer education. *Studies in Health Technology and Informatics*, 310, 1246-1250.