Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/115/2025.18295

Machine Learning-Based Research on the Adaptability of Adolescents to Online Education

Mingwei Wang[1,a,*], Xulingwei Shen[2,b], Sitong Liu[3,c]

[1]College of Humanities & Social Sciences, HZAU, Wuhan, China
[2]Shenghua Zizhu Academy, Shanghai, China
[3]GSE, University of Pennsylvania, Philadelphia, USA
a. 1683485281@qq.com, b. 775705937@qq.com, c. 1st9ts@gmail.com
**corresponding author*

Abstract: With the rapid advancement of internet technology, the adaptability of adolescents to online learning has emerged as a focal point of interest within the educational sphere. However, the academic community's efforts to develop predictive models for adolescent online learning adaptability require further refinement and expansion. Utilizing data from the "Chinese Adolescent Online Education Survey" spanning the years 2014 to 2016, this study implements five machine learning algorithms—logistic regression, K-nearest neighbors, random forest, XGBoost, and CatBoost—to analyze the factors influencing adolescent online learning adaptability and to determine the model best suited for prediction. The research reveals that the duration of courses, the financial status of the family, and age are the primary factors affecting students' adaptability in online learning environments. Additionally, age significantly impacts students' adaptive capacities. Among the predictive models, the random forest, XGBoost, and CatBoost algorithms demonstrate superior forecasting capabilities, with the random forest model being particularly adept at capturing the characteristics of students' adaptability.

Keywords: Online Learning Adaptability, Machine Learning, Adolescent.

1. Introduction

In the digital era, the evolution of information technology has catalyzed significant innovation in educational paradigms, with online learning emerging as a pivotal segment within the field of education[1]. The 50th Statistical Report on Internet Development in China, published in 2022, indicates a substantial growth in the user base of online education in China, expanding from 110 million in 2015 to 377 million by 2022. Despite the evident advantages of online education, the industry confronts numerous challenges[2].

Adaptability to online learning is the ability to adjust to personal and environmental changes during internet-based knowledge acquisition, aligning with the learning context to achieve goals [3]. Despite extensive research on factors like individual, familial, and scholastic dimensions, there is limited analysis of their interactions or predictive modeling. Most studies rely on descriptive statistics, hindering deeper insights into trends in adolescent adaptability. Developing predictive models to assess adaptability is essential for advancing online education understanding and practice.

DOI: 10.54254/2755-2721/115/2025.18295

This study investigates factors influencing adolescents' adaptability to online learning and develops predictive models to inform educational practice and policy. It refines the concept of online learning adaptability and examines dimensions such as personal background, educational environment, and learning conditions. Using a dataset with 13 features and one target column, univariate analysis explores feature distributions, while multivariate analysis uncovers relationships, highlighting the effects of gender and academic majors on adaptability.

Five machine learning algorithms—logistic regression, K-nearest neighbors, random forest, XGBoost, and CatBoost—are employed to establish and evaluate predictive models. The study identifies the most effective model for predicting adaptability and ensures its robustness through optimization. These findings enhance theoretical understanding and provide practical guidance for educators and policymakers, supporting personalized educational services, improving quality, and promoting equity in online learning.

2. Related work

Adolescents, as key beneficiaries of online education, experience changes in learning environments and strategies that impact outcomes. These shifts involve cognitive processes and knowledge acquisition, emphasizing the importance of adaptability. Key factors influencing adaptability can be categorized into four dimensions: individual, familial, scholastic, and community, guiding predictions of receptivity to online learning.

Individual factors, such as a student's ethnicity[4], are closely related to online learning adaptability; personal attributes, particularly psychological health status[5] and self-efficacy[6], significantly influence students' adaptability to online learning. Regarding the scholastic environment, the school setting plays a pivotal role in adolescents' adaptation to online education. The quality of teaching and course information[7] indirectly affects learning adaptability, exerting a notable positive influence. At the familial level, the accumulation of parental social, cultural, and economic[8] capital fosters students' adaptability to online learning. Additionally, the family provides a robust social support system, thereby enhancing students' adaptability and academic resilience in online learning[9]. The nature of the community in which students reside also impacts their adaptability to online learning, primarily manifesting as differences between urban and rural communities[10].

Current research on adolescents' adaptability to online learning explores individual, family, school, and community dimensions but lacks analysis of their interactions and systematic theoretical guidance, limiting applicability to educational practice. Many studies adopt single-disciplinary perspectives or rely on descriptive statistics, restricting predictive insights into adolescents' adaptability trends.

This paper addresses these gaps using social ecological systems theory, integrating perspectives from education, data science, and sociology. It systematically analyzes feature distributions through univariate analysis and explores complex interrelationships using multivariate analysis. By employing predictive modeling, this study identifies the most effective model for forecasting adolescents' adaptability to online learning. This approach provides a comprehensive framework, enhancing theoretical understanding and offering practical guidance for personalized, data-driven educational strategies.

3. Data Collection and Feature Analysis

3.1. Feature Description

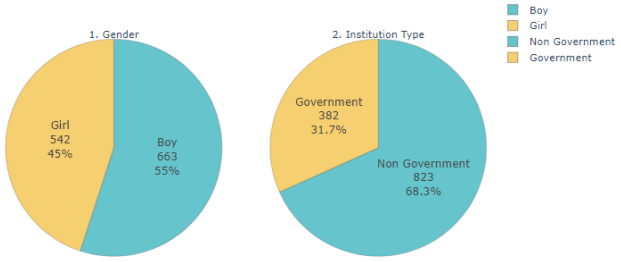
Features are attributes describing data samples, serving as input for machine learning models to predict or classify outcomes. This study uses features like gender, age, educational level, and internet type to represent students' backgrounds, educational settings, and online learning circumstances.

DOI: 10.54254/2755-2721/115/2025.18295

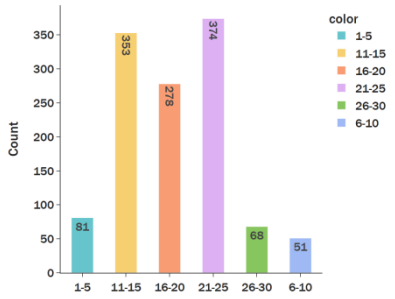
Feature selection is crucial for improving model accuracy and generalizability.The target variable, representing students' online learning adaptability, serves as the dependent variable the model predicts or classifies. By analyzing the relationship between features and adaptability, the model provides insights and recommendations for educators and policymakers to enhance online education strategies.

3.2. Univariate Analysis Results

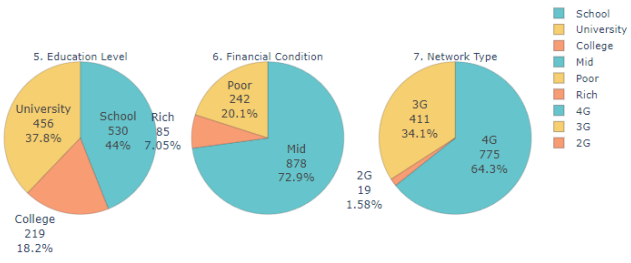
In the univariate analysis, the gender feature reveals that males and females constitute 55% and 45% of the total population, respectively. Regarding the type of educational institution, non-governmental institutions account for a significant proportion of 68.3%. The age distribution of respondents, mainly between 11 and 25 years old, reflects a general trend in the age distribution of the surveyed population. The class duration predominantly ranges from 1 to 3 hours, indicating a preference and acceptance level for course duration among respondents. The majority of respondents are engaged in higher education levels at schools and universities. In terms of financial condition, most respondents are in a moderate financial state, potentially influencing their degree and manner of participation in online education. The use of mobile data and phones is notably high at 57.7% and 84.1%, respectively, underscoring the widespread application of mobile networks and devices in online education. The adaptability level shows that approximately 51.9% of respondents indicate a moderate level of adaptability, 39.8% a low level, and the remainder a high level, providing crucial insights into the adaptability of the surveyed individuals. Collectively, these univariate analysis results offer vital clues and perspectives for an in-depth understanding of respondents' characteristics and behaviors.



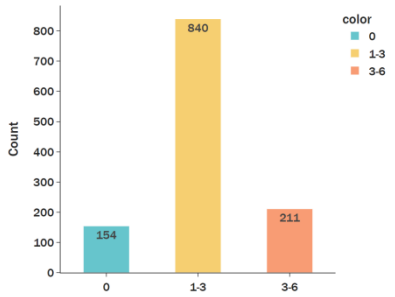
Gender & Institution Type



Age



Education Level & Financial Condition & Network Type



Class Duration

DOI: 10.54254/2755-2721/115/2025.18295

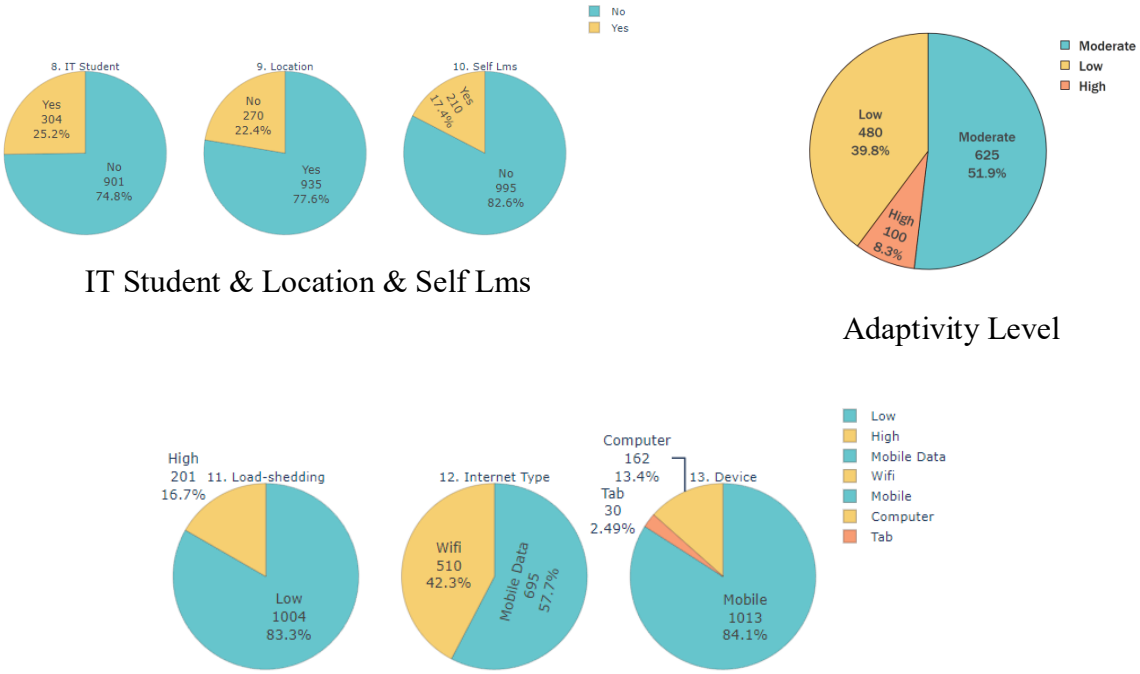


Figure 1: Visualization of Univariate Analysis Results from the Dataset

4. Model Establishment and Evaluation

4.1. Data Preprocessing

During the data preprocessing phase, features often require encoding to convert categorical or textual data into a numerical form that models can process. In this case, label encoding was utilized to transform the categorical values of each feature into numerical codes. This ensures that the model can correctly comprehend the relationships between features and process them appropriately.

To train and evaluate the machine learning models, the dataset is typically divided into training and testing sets. The training set is used to train the model, while the testing set is employed to assess model performance. Here, the dataset was split into training and testing sets, with 70% of the data allocated for training, encompassing 843 samples, and 30% for testing, comprising 362 samples. The partitioning process ensures that the model has sufficient data for effective learning and evaluation during both training and testing.

4.2. Model Selection

This study employs five machine learning algorithms to construct and evaluate an online education adaptability prediction model: logistic regression, K-nearest neighbors, random forest, XGBoost, and CatBoost.

Logistic regression is a simple, interpretable algorithm for binary classification, suitable for large-scale data. K-nearest neighbors (KNN) uses instance-based learning, classifying data based on the labels of the nearest neighbors without prior model training, offering flexibility.

Random forest, an ensemble method, builds multiple decision trees to improve accuracy and robustness, handling missing data and outliers effectively. XGBoost, a gradient-boosted tree algorithm, iteratively refines predictions, excelling with structured data and high-dimensional features. CatBoost, another gradient-boosted algorithm, automatically handles categorical features and missing values, demonstrating robustness for large-scale and sparse data.

DOI: 10.54254/2755-2721/115/2025.18295

By comparing these algorithms, this study identifies the most suitable model for predicting students' adaptability to online education, ensuring a comprehensive evaluation of performance across diverse methods.

4.3. Cross-Validation Results and Comparison of Models

In this study, the five selected machine learning models underwent 5-fold cross-validation, and their average accuracy on the validation set was calculated. Figure 2 illustrates the accuracy scores of the classifiers. The results indicate that the random forest model performed best in cross-validation, achieving an accuracy of 0.896, followed by XGBoost and CatBoost models with 0.891 and 0.886, respectively, while the logistic regression model showed relatively poor performance at 0.688. This suggests that the random forest, XGBoost, and CatBoost models may be more appropriate for constructing the online education adaptability prediction model.The random forest model demonstrates excellent performance in handling high-dimensional features and large-scale data, while XGBoost and CatBoost models excel in capturing complex relationships in classification problems, making them effective for prediction. Therefore, when selecting the final model, we will focus on the performance and applicability of these three models and further optimize and fine-tune them to ensure robustness and generalizability.

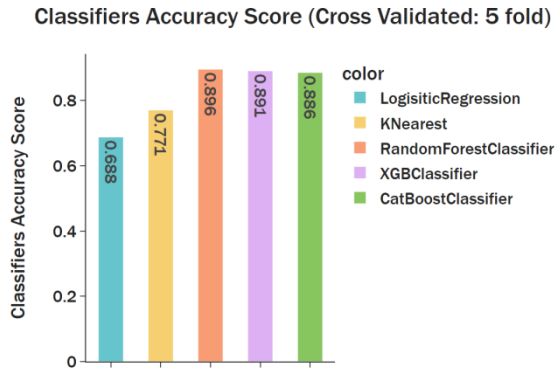


Figure 2: Classifiers' Accuracy Scores

5. Discussion and Analysis of Results

5.1. Analysis of the Impact of Features on Adaptability

Figure 3 presents the ranking of feature importance based on the random forest model, allowing for an analysis of the influence of various features on adaptability. The results reveal that "Class Duration," "Financial Condition," and "Age" have the most significant impact on adaptability. This suggests that the length of courses, the financial status of families, and the age of students are likely key determinants of students' adaptability in online education. Additionally, features such as gender, type of school, and type of network also have some influence on adaptability, albeit to a lesser extent.

DOI: 10.54254/2755-2721/115/2025.18295

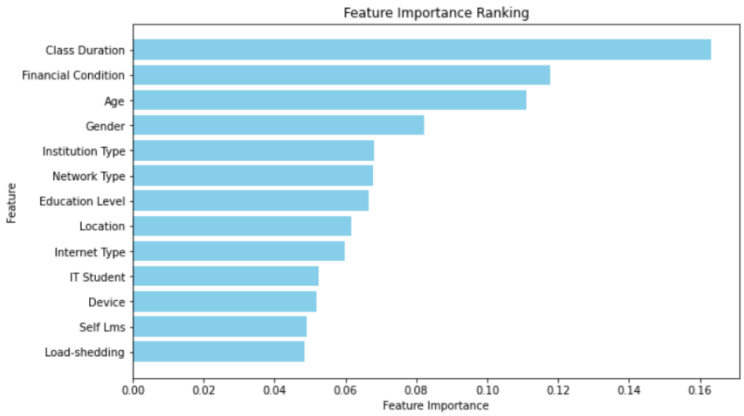


Figure 3: Ranking of Feature Importance

6. Conclusion and Prospects

This study aimed to gain an in-depth understanding of students' adaptability in online education and the factors influencing it through data analysis and machine learning model predictions. By analyzing and modeling the features in the dataset, we have reached several important conclusions:

The duration of courses, the financial status of families, and age are critical factors determining students' adaptability in online training. Longer course durations may require students to possess greater self-management skills and endurance, while better financial conditions provide students with superior learning conditions and resource support, conducive to enhancing their online learning performance and adaptability. Furthermore, age also affects students' adaptability, with older students likely having richer learning experiences and self-management skills, making them more adept at adapting to online learning environments.

Comparing the performance of different machine learning models reveals that the random forest, XGBoost, and CatBoost models perform well in predicting students.

References

[1] *Wu Wentao, Liu Hehai, Bai Qian. Building a Learning Society: Practicing Chinese Modernization through Educational Digitization [J]. Chinese Journal of Educational Technology, 2023, (03): 17-24+45.*

[2] *Zhou Hongwei. Research on the Adaptive Learning Model of Online Education Based on Educational Big Data and Its Application [J]. Research on Continuing Education, 2023, (03): 110-114.*

[3] *Tian Lan. Review of Research on Learning Adaptability of Primary and Secondary School Students in China [J]. Psychological Science, 2004, (02): 502-504. DOI: 10.16719/j.cnki.1671-6981.2004.02.077.*

[4] *Chi Shangxin, Kang Shihan, Zeng Wenqian, et al. Research on College Students' Online Learning Adaptability and Influencing Factors in the Digital Age—From the Perspective of Ecological Systems Theory [J]. Chinese Journal of Youth and Social Sciences, 2023, 42(06): 73-87. DOI: 10.16034/j.cnki.10-1318/c.2023.06.012.*

[5] *Alqurashi, E. (2016). Self-efficacy in online learning environments: A literature review. Contemporary Issues in Education Research (Online), 9(1), 45.*

[6] *Shen Zhonghua, Wu Daguang. An Investigation of the Influencing Factors of College Students' Online Learning Effectiveness and Satisfaction: An Empirical Analysis Based on the Structural Equation Model [J]. Research in Educational Development, 2020, 40(11): 25-36+59. DOI: 10.14121/j.cnki.1008-3855.2020.11.006.*

[7] *Li Fengling. Research on the Influencing Factors of Online Learning Adaptability of Vocational College Students during the Epidemic [D]. Guangdong University of Technology, 2022. DOI: 10.27729/d.cnki.ggdjs.2022.000073.*

[8] *Wang Weidong, Dong Tianyi. Research on the Adaptability of Junior High School Students to Online Learning at Home—Based on an Empirical Analysis of 60,160 Questionnaires [J]. Educational Guidance, 2024, (04): 53-65. DOI: 10.16215/j.cnki.cn44-1371/g4.2024.04.002.*

[9] *Permatasari, N., Ashari, F. R., & Ismail, N. (2021). Contribution of perceived social support (peer, family, and teacher) to academic resilience during COVID-19. Golden Ratio of Social Science and Education, 1(1), 01-12.*

DOI: 10.54254/2755-2721/115/2025.18295

[10] Chen Tingyu, Yang Yanqing. Research on the Current Status and Influencing Factors of Online Learning Adaptability of Nursing Major Students in Vocational Colleges [J]. Health Vocational Education, 2024, 42(14): 80-83. DOI: 10.20037/j.issn.1671-1246.2024.14.22.