



ISSN: 2723-9535

Available online at [www.HighTechJournal.org](http://www.HighTechJournal.org)

# HighTech and Innovation Journal

Vol. 5, No. 2, June, 2024



## Analysis of Factors Influencing Online Learning Using the Decision Tree Method

Xiaojie Li <sup>1\*</sup>, Huili Tang <sup>1</sup>

<sup>1</sup> Department of Pre-Primary Education, Zhengzhou Preschool Education College, Henan 450000, China.

Received 03 December 2023; Revised 22 April 2024; Accepted 04 May 2024; Published 01 June 2024

### Abstract

**Objective:** With the continuous development of online learning, the analysis of students' online learning has become increasingly important. Understanding which factors can influence students' engagement in online learning plays a crucial role in improving their learning performance. **Methods:** By utilizing web crawling techniques, students' online learning behavior data was collected from the Chinese University's massive open online courses (MOOC) platform. To address the imbalance in the dataset, a synthetic minority oversampling technique (SMOTE) was used. Course progress was used to reflect students' online learning status, which was categorized into interruptions and completions. Furthermore, to tackle the issue of low computational efficiency in the C4.5 decision tree algorithm, its calculation formula was improved to develop an improved version of C4.5. **Findings:** Of the several factors analyzed, the number of course chapters had the greatest impact on students' online learning, followed by the number of course evaluations and overall course scores. The classification of students' online learning situations based on an improved C4.5 algorithm revealed that the improved method achieved the highest accuracy rate of 0.942 and the shortest classification time of 0.165 s compared to methods such as the naive Bayesian and random forest algorithms. **Novelty:** This study designed an improved version of C4.5 to analyze the influencing factors in online learning, and its reliability was demonstrated through experiments, providing a new effective method for data analysis in online learning.

**Keywords:** Higher Education; Decision Tree; Online Learning; Influence Factors; e-Learning.

## 1. Introduction

With the development of information technology, the traditional teaching model has seen continuous changes, and the emergence of massive open online courses (MOOCs) [1] has provided more students with the possibility of online learning. Compared with the traditional teaching mode, MOOCs have the advantages of high freedom, high openness, and low learning costs [2], so they have been widely used, but interrupting online learning is frequent [3], i.e., the enrollment rate of students is high but the completion rate of learning is low. In order to further enhance the effectiveness of students' online learning, it is necessary to analyze the influencing factors. Students have accumulated massive learning-related behavioral data through online learning, which is more convenient to obtain and process. With the advancement of technologies such as data mining, more and more methods have been applied in educational data processing [4]. Dhanalakshmi et al. [5] used the Apriori algorithm to predict the performance of special children (e.g., mentally retarded) in special school learning, thus helping teachers to nurture the children. Sang [6] recognized college students' psychological crisis states based on data mining and found through experiments that the method obtained an accuracy of more than 90%, which was helpful to help psychological management personnel guide their students.

\* Corresponding author: [fpjx41781@sina.cn](mailto:fpjx41781@sina.cn)

 <http://dx.doi.org/10.28991/HIJ-2024-05-02-018>

➤ This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights.

Thangakumar et al. [7] designed a student achievement classification method based on ant colony optimization and logistic regression and found through experiments that its highest accuracy reached 97.99% and the F1 value was 97.02%. Roslan et al. [8] predicted the dropout rate of college students through the decision tree and regression model, conducted experiments on student data from a private university in Malaysia, and found that the decision tree method obtained a classification performance of 89.49%, i.e., the method was effective in predicting the at-risk students accurately. Ulkhaq et al. [9] proposed the term 'university bias' to address judgment biases in student competitions, as judges may give higher scores to participants from the same university. They conducted an analysis of a dataset of Indonesia's annual national university student competition using association rule mining techniques and found the existence of university bias. Onyema et al. [10] studied the influence of mobile technology on physical education during the COVID-19 lockdown period and evaluated its effects on academic performance among teachers. Using regression analysis and analysis of variance, they found that the application of mobile technology had a statistically significant impact on academic performance for both teachers and students, and the assistance provided by mobile technology was particularly valuable for continuing education.

Bessadok et al. [11] clustered student activities using activity log attributes from a learning management system dataset and subsequently examined the correlation between the profiles and academic achievements through statistical analysis, providing evidence for the influence of students' personal profiles on their educational performance. In another study by Tian et al. [12], they constructed a learning interest classification model based on a text convolutional neural network and gated recurrent unit. They evaluated its effectiveness using an experimental dataset comprising online English learners and found that it was excellent at classifying learning interests.

Pham et al. [13] developed a non-spatiotemporal multidimensional model for assessing the quality of university education by using the features of the dataset structure to form data variables and analyzing the relationships between these variables. By employing this approach, it enabled university administrators to formulate policies and enhance educational quality effectively. The article first analyzed the influencing factors of students' online learning in Section 1 and used the synthetic minority oversampling technique (SMOTE) to balance the dataset. Then, in Section 2, the C4.5 algorithm was analyzed, and its calculation formula was improved. In Section 3, the improved C4.5 method was utilized to examine the influencing factors of students' online learning, and its classification effectiveness and time were compared with the other approaches. Finally, Section 4 concludes the whole article. The workflow diagram is provided in Figure 1.

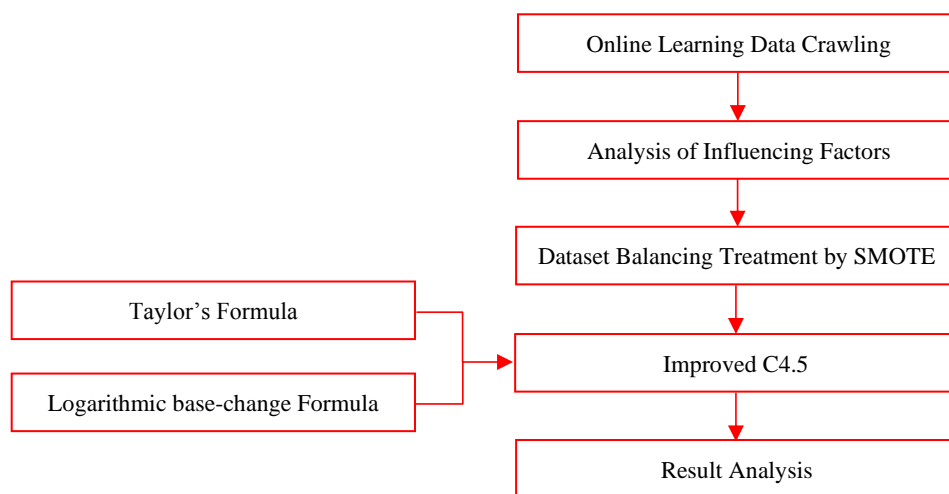


Figure 1. The workflow diagram

## 2. Factors Influencing Students to Learn Online

Online learning is a way of learning with the help of information technology, which has become a very important part of modern education. It can get rid of the limitations of time and space [14] and promote the development of independent learning and lifelong learning. Online learning can help students learn the content of many different subjects more conveniently and efficiently. Compared with traditional learning methods, online learning has the following characteristics: (1) diversified learning methods: students can learn freely through personal computers and mobile devices; (2) personalized learning content: online learning can provide students with more targeted content to meet their personalized learning needs; (3) interactive learning communication: in the process of online learning, students can interact with their classmates and teachers through forum messages, scoring, and other ways.

In order to enhance students' online learning efficiency, it is necessary to analyze the factors that influence students' online learning. Therefore, this paper aims to establish a dataset using web scraping techniques and obtain relevant influencing factors related to students' online learning. The obtained dataset was balanced using the SMOTE algorithm.

Subsequently, an improved C4.5 algorithm was designed to understand the impact of different factors on whether students continue their online learning and predict in advance if they are likely to discontinue their studies. The specific research content is as follows:

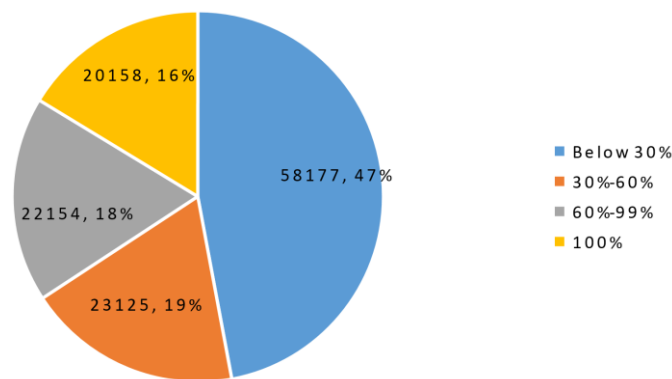
This paper adopts crawling technology to crawl students' behavioral data on online learning from the China University MOOC [15], and the target students were those who receive higher education, i.e., college, bachelor, master, and doctorate. A total of about 20,000 higher education students' behavioral data were crawled from October 10, 2020, to October 10, 2022. After cleaning the data and excluding incomplete and abnormal values, 123,614 behavioral data points were obtained.

To facilitate the subsequent decision tree analysis, the obtained influencing factors related to online learning were processed, as shown in Table 1.

**Table 1. Factors influencing students' online learning**

Influencing Factors	Description
Students' education background	College = 1, Bachelor = 2, Master = 3, Doctor = 4
Number of course participants	The number of learners shown on the course page, below 1000 = 0, above 1000 = 1
Overall course score	The overall score displayed on the course page, below 4.0 = 0, above 4.0 = 1
Number of course evaluations	The number of evaluations displayed on the course page, below 100 = 0, above 100 = 1
Number of course chapters	The number of chapters displayed on the course page, below 10 = 0, above 10 = 1
Number of course tests	The number of tests displayed on the course page, below 10 = 0, above 10 = 1
Course learning progress	The average of the progress of different courses in which the students are involved, below 100% = 0, 100% = 1

First, the average progress of students' courses was analyzed. The distribution of course learning progress among the 1236,145 data collected is shown in Figure 2.



**Figure 2. Distribution of students' course learning progress**

It was observed in Figure 2 that among the collected samples, the number of students whose learning progress in the course was below 30% was the largest, accounting for 47%; 19% of the students had a learning progress of 30%~60%, 18% of the students had a learning progress of 60%~99%, and 16% of the students had a learning progress of 100%. If the course learning progress was divided into two categories, namely, interrupted learning (learning progress not reached 100%) and completed learning (learning progress reached 100%), it was found that the distribution of samples in the two categories was relatively uneven, accounting for 84% (10,3456) and 16% (20,158); therefore, to avoid the impact of data imbalance on the results of decision tree analysis, this paper chose the SMOTE method [16] for sample processing.

The principle of SMOTE is as follows. It was assumed that there was original data set  $S$  and minority-category sample set  $Y = \{y_1, y_2, \dots, y_m\}$ . For every sample  $y_i$ ,  $K$  neighbors were obtained through the K-nearest neighbors' algorithm [17].  $s$  samples were randomly selected, and a new sample was generated for every  $y_j$  using the formula  $Y_{new} = Y_i + rand(0,1) \times (y_j - y_i)$ . New samples were constantly generated until the dataset was balanced. After SMOTE processing, the ratio of interrupted learning to completed learning samples reached 1:1, i.e., the number of samples was 103,456:103,456.

On this basis, the correlation between different factors and students' learning progress was analyzed using Pearson correlation coefficients, and the results are presented in Table 2.

**Table 2. The calculation results of correlation coefficients**

Influence factor	Correlation coefficient
Students' educational background	0.749
Number of course participants	0.612
Overall course score	0.856
Number of course evaluations	0.874
Number of course chapters	0.926
Number of course tests	0.684

According to Table 2, there is a strong correlation between the six selected factors and students' learning progress. Among them, the number of course chapters, the number of course evaluations, and the overall course score had high correlation coefficients exceeding 0.8, indicating a strong relationship with students' learning progress. Student education background, the number of course tests, and the number of course participants followed closely behind in terms of their correlation. The results from Table 2 demonstrated that these six selected factors could be used for further analysis.

### 3. Decision Tree Method

A decision tree is a top-down classification method with a relatively simple and understandable computational process, which has extensive applications in data processing [18]. ID3 is one of the classical decision tree methods [19], which splits nodes by information gain (Gain). For data set  $X = \{x_1, x_2, \dots, x_n\}$ , it is assumed that the occurrence probability of every data is  $p(i)$ , then, the information value of  $x_i$  is written as:  $I(x_i) = -\log_2 p(i)$ , and the information entropy of  $X$  is written as:  $Info_{(X)} = -\sum_{i=1}^n p(i) \log_2 p(i)$ . It is assumed that the data set is divided into  $C$  categories:  $C = \{c_1, c_2, \dots, c_n\}$ , the occurrence probability of every category is  $p(c_n)$ . Then, the information value of  $c_i$  is written as:  $I(c_i) = -\log_2 p(c_i)$ , and the information entropy of  $C$  is written as:  $Info_{(C)} = -\sum_{i=1}^n p(c_i) \log_2 p(c_i)$ .

The dataset is classified by feature  $A$ . The corresponding information gain of  $A$  is written as:

$$Gain_{(A)} = Info_{(X)} - Info_A(X) \tag{1}$$

ID3 determines the classification criteria of a decision tree based on the size of the gain value. C4.5 is improved on the basis of ID3 [20]. The information gain ratio is used as an indicator for classification. For feature  $A$ , its corresponding information gain ratio is:

$$GainRatio_{(A)} = \frac{Gain_{(A)}}{SplitInfo_A(X)}, \tag{2}$$

$$SplitInfo_A(X) = -\sum_{j=1}^m \frac{|X_{Aj}|}{|X|} \log_2 \frac{|X_{Aj}|}{|X|}, \tag{3}$$

where  $SplitInfo_A(X)$  is the splitting information of feature  $A$ .

C4.5 involves logarithmic operations in the calculation process, which leads to low calculation efficiency. To solve this problem, this paper improved the calculation formula of C4.5. According to Taylor's formula and logarithmic base change formula [21]:

$$\ln(1 + x) = \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{n} x^n, \tag{4}$$

$$lb x = \frac{\ln x}{\ln 2}. \tag{5}$$

The formula of  $GainRatio$  in C4.5 was improved as:

$$GainRatio(C, A) = \frac{Gain_{(C,A)}}{SplitInfo_A(A)} = \frac{\sum_{i=1}^n \frac{|X_{ci}| \times (|X| - |X_{ci}|)}{|X|} - \sum_{j=1}^m \sum_{i=1}^n \frac{|X_{ciAj}| \times (|X_j| - |X_{ciAj}|)}{|X_{Aj}|}}{\sum_{j=1}^m \frac{|X_{Aj}| \times (|X| - |X_{Aj}|)}{|X|}}. \tag{6}$$

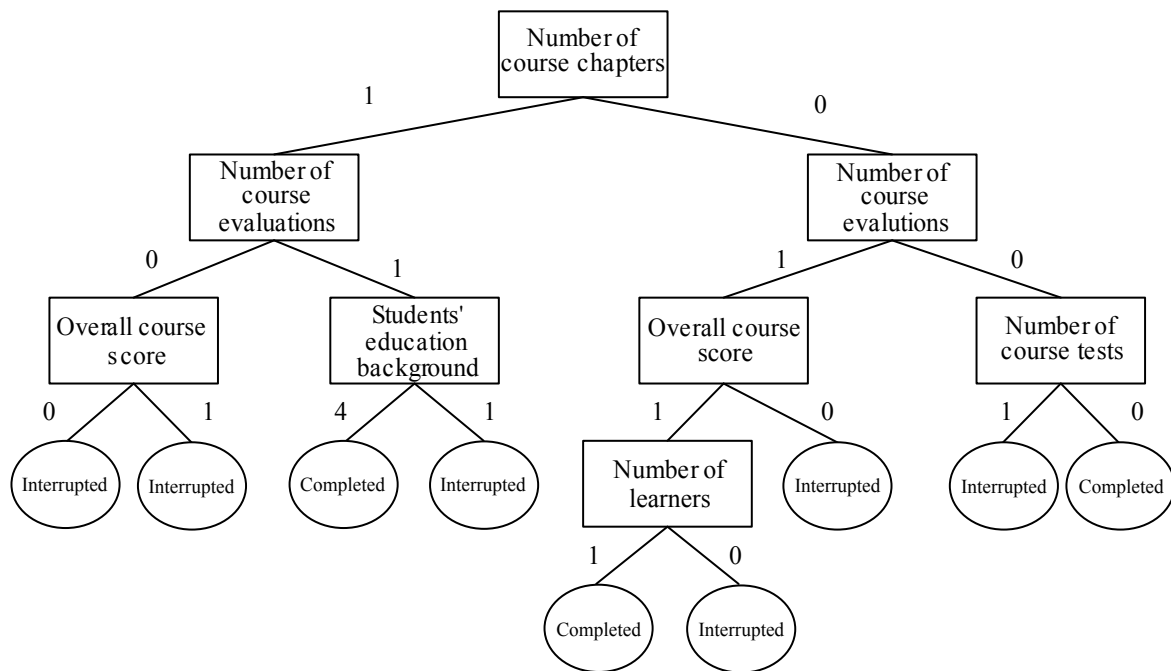
### 4. Results and analysis

The influencing factors of students' online learning were analyzed using the optimized C4.5 method. First, the ranking outcomes of the importance of the six influencing factors selected are shown in Table 3.

**Table 3. Ranking of importance of influencing factors**

Importance ranking	Influencing factors
1	Number of course chapters
2	Number of course evaluations
3	Overall course score
4	Students' education background
5	Number of course tests
6	Number of learners

It was seen from Table 3 that the number of course chapters had the greatest impact on students' performance in online learning, followed by the number of course evaluations, while the number of course tests and the number of learners had small impacts on students' performance of online learning. Therefore, the C4.5 decision tree obtained by using the number of course chapters as the root node is shown in Figure 3.



**Figure 3. The decision tree for analyzing the factors influencing students to learn online**

According to Figure 3, the following rules can be obtained.

- (1) IF number of course chapters = 1 AND number of course evaluations = 0 AND overall course score = 0 THEN learning progress = interrupted;
- (2) IF number of course chapters = 1 AND number of course evaluations = 0 AND overall course score = 1 THEN learning progress = interrupted;
- (3) IF number of course chapters=1 AND number of course evaluations=1 AND students' education background = 4 THEN learning progress = completed;
- (4) IF number of course chapters = 1 AND number of course evaluations = 1 AND students' education background = 1 THEN learning progress = interrupted;
- (5) IF number of course chapters = 0 AND number of course evaluations = 1 AND overall course score = 1 AND number of learners = 1 THEN learning progress = completed;
- (6) IF number of course chapters = 0 AND number of course evaluations = 1 AND overall course score = 1 AND number of learners = 0 THEN learning progress = interrupted;
- (7) IF number of course chapters = 0 AND number of course evaluations = 1 AND overall course score = 0 THEN learning progress = interrupted;
- (8) IF number of course chapters = 0 AND number of course evaluations = 0 AND number of course tests = 1 THEN learning progress = interrupted;
- (9) IF number of course chapters = 0 AND number of course evaluations = 0 AND number of course tests = 0 THEN learning progress = completed.

Based on the decision tree and the analysis of the above rules, it was found that the number of course chapters had the greatest impact on the course learning progress when students studied online. A large number of course chapters indicated that the time and effort required to finish the course was also high. Facing a long course in the process of online learning, higher education students might stop learning online due to a lack of time and patience, which led to interruption of learning progress. The number of course evaluations and the overall course score could reflect the quality and hotness of the course to a certain extent. The high number of evaluations and the high overall score indicated that the course was popular among students and the enthusiasm and initiative of the students were high; therefore, the possibility of students completing the study was high when the number of course evaluations was high and the score was high. Students' educational background, number of course tests, and number of learners had small influences on online learning. Students were more likely to discontinue learning when they had a low education background, the number of course tests were large, and few learners studied the course.

Based on the results of the analysis, it was concluded that teachers should reasonably allocate course chapters to try to avoid students interrupting their learning due to the long course, and at the same time, they should improve the quality of the courses, fully mobilize students' learning initiative and enthusiasm, and strengthen students' willingness to learn online to motivate them to complete their online learning.

In the current research on the factors influencing students to engage in online learning, Effendy et al. [22] found that students' intention to use online learning is influenced by the quality of instruction; Elshami et al. [23] found that technopedagogical skills are the most important factor influencing the online learning of students; Syafril et al. [24] found that the most important barrier for students to engage in online learning is the lack of preparation of online learning materials. However, in the preliminary data investigation, most of these studies were conducted through questionnaire surveys. In comparison to these studies, this article first achieves higher objectivity in the data collection stage by obtaining and analyzing data from students' online behaviors. Therefore, the obtained influencing factors were also more reliable, providing stronger support for practical online learning.

Finally, in order to understand the classification performance of the C4.5 model for online learning situations, the classification performance was analyzed. The results of the model classification are shown in Figure 4.

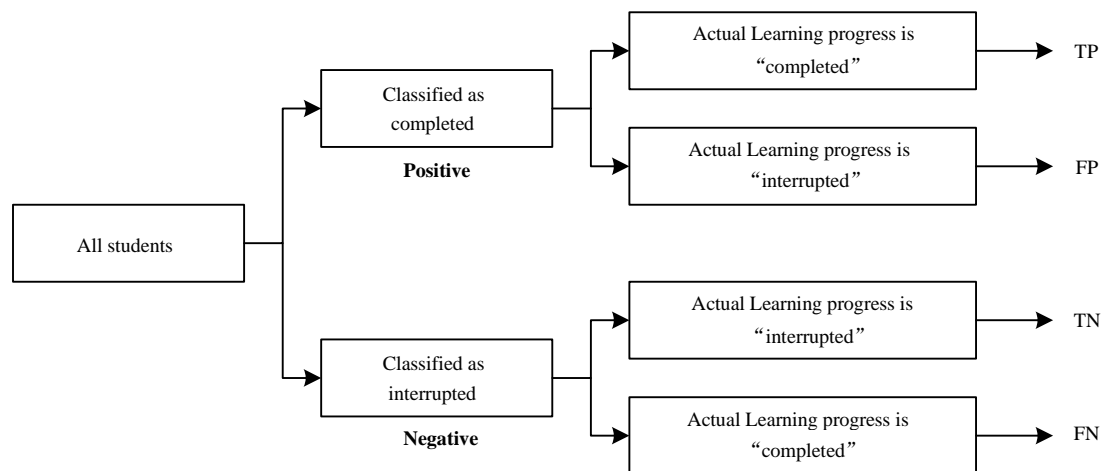


Figure 4. Model classification results

The evaluation indicators of classification effectiveness are as follows:

$$(1) Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

$$(2) Precision = \frac{TP}{TP+FP}$$

$$(3) Recall = \frac{TP}{TP+FN}$$

$$(4) F1 = \frac{2*Precision*Recall}{Precision+Recall}$$

To understand the effectiveness of the improved C4.5 method, it was compared with the baseline methods, including naive Bayesian (NB) [25], random forest (RF) [26], and support vector machine (SVM) methods [27], and the results are shown in Figure 5.

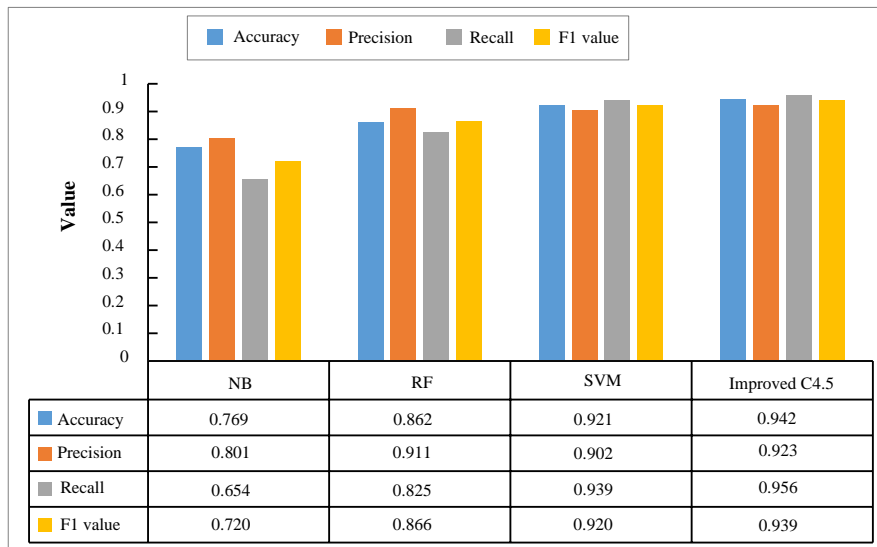


Figure 5. Comparison of the classification effects between different methods

It was seen from Figure 5 that the NB method had poor performance in accuracy and recall rate and had an F1 value of 0.720, indicating that the method was not effective in classifying students' learning progress in the course. The indicators of the RF method were above 80%, among which, the accuracy reached 0.911. The indicators of the SVM method were above 90%, indicating that the SVM method performed better than the former two methods, and the F1 value of the SVM method was 0.920, which was 0.2 larger than the NB method. Finally, the improved C4.5 method achieved 0.942 in accuracy, 0.923 in precision, and 0.956 in recall rate in classifying students' course learning progress. The F1 value of the improved C4.5 method was 0.939, which was 0.219 larger than the NB method, 0.073 larger than the RF method, and 0.019 larger than the SVM method. The results verified that the improved C4.5 method was reliable in classifying students' online learning situations and predicting whether they will interrupt or complete their learning.

Then, the time performance of different methods was compared. Since the calculation formula of the C4.5 method was improved, the traditional C4.5 method was also added to the comparison, and the results are presented in Figure 6.

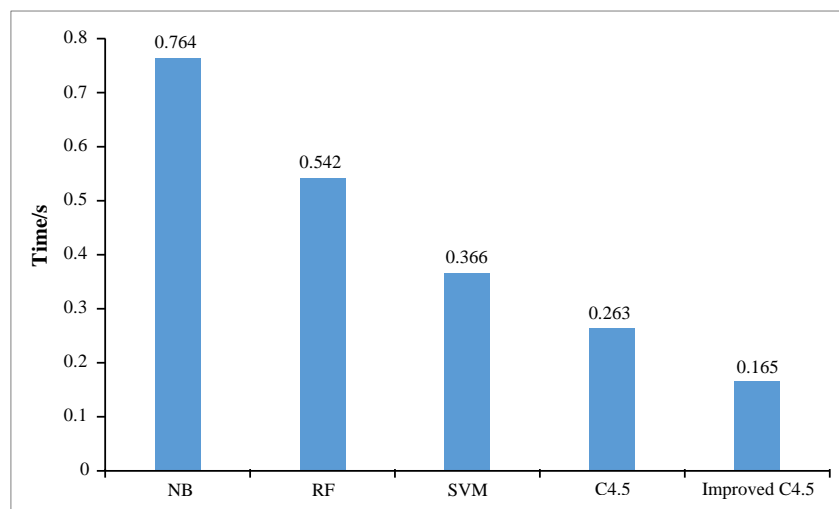


Figure 6. Comparison of the time performance between different methods

It was seen from Figure 6 that, among the compared methods, the classification time required by the C4.5 method was significantly shorter. The classification times of the NB, RF, and SVM methods were 0.764 s, 0.542 s, and 0.366 s, respectively. The classification time required by the C4.5 method was 0.263 s, which was significantly shorter than that of the former methods. After the improvement of the calculation formula, the classification time required by the improved C4.5 method was 0.165 s, which was 0.098 s shorter than the traditional C4.5 method. These results verified that the improvement of the C4.5 method significantly shortened the classification time and thus improved the efficiency of classification. Overall, the improved C4.5 method not only performed well in classifying students' online student profiles but also had good time performance, so the method can be further applied in practical online learning.

Based on the experimental results, the improved C4.5 method identified the factors that have a great influence on students' online learning situations. In practical application, these results can provide some suggestions for the online

learning platform, teachers, and even students. For the online learning platform, further analysis can be conducted on the courses with high and low student completion in order to improve the attractiveness of the platform for students. Teachers can consider improving their technical skills to arouse students' interest in learning, and students can choose courses with a moderate number of chapters, more evaluations, and higher scores to achieve better learning results when they do online learning. The performance comparison of the improved C4.5 method with other methods also further proves its reliability, good classification effect, and higher classification efficiency, so it has good practical value.

## 5. Conclusion

This study analyzed the influencing factors of online learning for higher education students through the decision tree method. An improved C4.5 method was designed to enhance computational efficiency. Through analysis, it was found that the number of course chapters and the number of course evaluations had a significant impact on students' learning progress. The decision tree model developed achieved an accuracy rate of 0.942 in classification experiments, with a classification time of only 0.165 s. These results outperformed the other methods, such as NB and RF, demonstrating the reliability of the proposed approach in classifying students' online learning situations.

The research findings of this article are beneficial in helping educators further understand the influencing factors of students' online learning and provide a theoretical basis for the design of practical online courses. It can contribute to improvements in the planning and design of online courses to enhance students' motivation and drive to complete study, thereby improving the efficiency and quality of online learning and promoting its further development.

However, this study has some limitations. For example, the source of data was relatively single, and the selection of influencing factors was not comprehensive enough. Therefore, in future work, the researchers will conduct research on more comprehensive and complex student behavior data and include more influencing factors to further validate the applicability of the proposed method.

## 6. Declarations

### 6.1. Author Contributions

Conceptualization, X.L. and H.T.; methodology, X.L.; software, H.T.; validation, X.L. and H.T.; formal analysis, H.T.; data curation, X.L.; writing—original draft preparation, X.L.; writing—review and editing, X.L.; visualization, H.T.; supervision, H.T.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

### 6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### 6.4. Institutional Review Board Statement

Not applicable.

### 6.5. Informed Consent Statement

Not applicable.

### 6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## 7. References

- [1] Feitosa de Moura, V., Alexandre de Souza, C., & Noronha Viana, A. B. (2021). The use of Massive Open Online Courses (MOOCs) in blended learning courses and the functional value perceived by students. *Computers and Education*, 161, 1–14. doi:10.1016/j.compedu.2020.104077.
- [2] Khalid, A., Lundqvist, K., & Yates, A. (2022). A literature review of implemented recommendation techniques used in Massive Open online Courses. *Expert Systems with Applications*, 187, 1–16. doi:10.1016/j.eswa.2021.115926.
- [3] Goli, A., Chintagunta, P. K., & Sriram, S. (2022). Effects of Payment on User Engagement in Online Courses. *Journal of Marketing Research*, 59(1), 11–34. doi:10.1177/00222437211016360.



- [4] Yahya, A. A., & Osman, A. (2019). A data-mining-based approach to informed decision-making in engineering education. *Computer Applications in Engineering Education*, 27(6), 1402–1418. doi:10.1002/cae.22158.
- [5] Dhanalakshmi, R., Muthukumar, B., & Canessane, R. A. (2022). Analysis of Special Children Education Using Data Mining Approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 30, 125–140. doi:10.1142/S0218488522400074.
- [6] Sang, H. (2021). Analysis and Research of Psychological Education Based on Data Mining Technology. *Security and Communication Networks*, 2021, 1–8. doi:10.1155/2021/8979507.
- [7] Thangakumar, J., & Kommina, S. B. (2020). Ant colony optimization-based feature subset selection with logistic regression classification model for education data mining. *International Journal of Advanced Science and Technology*, 29(3), 5821–5834.
- [8] Et.al, N. R. (2021). Prediction of Student Dropout in Malaysian's Private Higher Education Institute using Data Mining Application. *Turkish Journal of Computer and Mathematics Education*, 12(3), 2326–2334. doi:10.17762/turcomat.v12i3.1219.
- [9] Ulkhaq, M. M., Pramono, S. N. W., & Adyatama, A. (2023). Assessing the tendency of judging bias in student competition: a data mining approach. *Journal of Applied Research in Higher Education*, 15(4), 1198–1211. doi:10.1108/JARHE-02-2022-0053.
- [10] Onyema, E. M., Khan, R., Eucheria, N. C., & Kumar, T. (2023). Impact of Mobile Technology and Use of Big Data in Physics Education During Coronavirus Lockdown. *Big Data Mining and Analytics*, 6(3), 381–389. doi:10.26599/BDMA.2022.9020013.
- [11] Bessadok, A., Abouzinadah, E., & Rabie, O. (2023). Exploring students' digital activities and performances through their activities logged in learning management system using educational data mining approach. *Interactive Technology and Smart Education*, 20(1), 58–72. doi:10.1108/ITSE-08-2021-0148.
- [12] Tian, X., Alassafi, M. O., & Alsaadi, F. E. (2023). An efficient english teaching driven by enterprise-social media big data: A neural network-based solution. *Fractals*, 31(6), 1–9. doi:10.1142/S0218348X23401515.
- [13] Pham, D. Van, & Nguyen, B. K. (2023). A Visual Analytics Approach Applying for Discovering Knowledge from Multivariate Datasets of Stakeholders Feedback in the University. *Vietnam Journal of Computer Science*, 10(4), 463–483. doi:10.1142/S2196888823500082.
- [14] Mubarak, A. A., Cao, H., & Hezam, I. M. (2021). Deep analytic model for student dropout prediction in massive open online courses. *Computers and Electrical Engineering*, 93, 1–14. doi:10.1016/j.compeleceng.2021.107271.
- [15] Icourse163 (2024). iCourse\_Free Online Course, Beijing, China. Available online: <https://www.icourse163.org/#> (accessed on March 2024).
- [16] Arwathananukul, S., Saengrayap, R., Chaiwong, S., & Aunsri, N. (2022). Fast and Efficient Cavendish Banana Grade Classification using Random Forest Classifier with Synthetic Minority Oversampling Technique. *IAENG International Journal of Computer Science*, 49(1), 46–54.
- [17] Gallego, A. J., Calvo-Zaragoza, J., Valero-Mas, J. J., & Rico-Juan, J. R. (2018). Clustering-based k-nearest neighbor classification for large-scale data with neural codes representation. *Pattern Recognition*, 74, 531–543. doi:10.1016/j.patcog.2017.09.038.
- [18] Lashgari, M., & Shahrtash, S. M. (2022). Fast online decision tree-based scheme for predicting transient and short-term voltage stability status and determining driving force of instability. *International Journal of Electrical Power and Energy Systems*, 137, 1–13. doi:10.1016/j.ijepes.2021.107738.
- [19] Es-Sabery, F., Es-Sabery, K., Qadir, J., Sainz-De-Abajo, B., Hair, A., García-Zapirain, B., & De La Torre-Díez, I. (2021). A MapReduce Opinion Mining for COVID-19-Related Tweets Classification Using Enhanced ID3 Decision Tree Classifier. *IEEE Access*, 9, 58706–58739. doi:10.1109/ACCESS.2021.3073215.
- [20] Irawan, Y. (2021). Application of the C4.5 Decision Tree Algorithm to Predict the Eligibility of Prospective Donors to Donate Blood Using Data Mining Classification. *JTIM: Journal of Information Technology and Multimedia*, 2(4), 181–189. doi:10.35746/jtim.v2i4.75.
- [21] Navarro, J. D., Millwater, H. R., Montoya, A., & Restrepo, D. (2022). Arbitrary-Order Sensitivity Analysis in Phononic Metamaterials Using the Multicomplex Taylor Series Expansion Method Coupled with Bloch's Theorem. *Journal of Applied Mechanics, Transactions ASME*, 89(2), 1–15. doi:10.1115/1.4052830.
- [22] Effendy, F., Kurniawati, O. D., & Priambada, G. (2021). Factor Affecting E-Learning User Acceptance: A Case Study of AULA. *Journal of Physics: Conference Series*, 1783(1), 1–6. doi:10.1088/1742-6596/1783/1/012122.
- [23] Elshami, W., Taha, M. H., Abdalla, M. E., Abuzaid, M., Saravanan, C., & Al Kawas, S. (2022). Factors that affect student engagement in online learning in health professions education. *Nurse Education Today*, 110, 105261. doi:10.1016/j.nedt.2021.105261.

- [24] Syafriil, & Novrianti. (2021). The Factor Analysis of Online Learning Barriers on Learning Evaluation Course during Covid-19 Pandemic. *Journal of Physics: Conference Series*, 1779(1), 1–5. doi:10.1088/1742-6596/1779/1/012021.
- [25] Mehmood, T., Kanwal, A., & Butt, M. M. (2022). Naive Bayes combined with partial least squares for classification of high dimensional microarray data. *Chemometrics and Intelligent Laboratory Systems*, 222, 1–8. doi:10.1016/j.chemolab.2022.104492.
- [26] Lu, M., Sadiq, S., Feaster, D. J., & Ishwaran, H. (2018). Estimating Individual Treatment Effect in Observational Data Using Random Forest Methods. *Journal of Computational and Graphical Statistics*, 27(1), 209–219. doi:10.1080/10618600.2017.1356325.
- [27] Marston, Z. P. D., Cira, T. M., Knight, J. F., Mulla, D., Alves, T. M., Hodgson, E. W., Ribeiro, A. V., MacRae, I. V., & Koch, R. L. (2022). Linear Support Vector Machine Classification of Plant Stress from Soybean Aphid (Hemiptera: Aphididae) Using Hyperspectral Reflectance. *Journal of Economic Entomology*, 115(5), 1557–1563. doi:10.1093/jee/toac077.