Academic Performance Prediction Method of Online Education using Random Forest Algorithm and Artificial Intelligence Methods

https://doi.org/10.3991/ijet.v16i05.20297

Jing Yu
Zhejiang Business College, Hangzhou, China
yj@zjbc.edu.cn

Abstract—In order to improve the teaching quality of online education, the prediction method of students' online academic performance has been studied. First, the learning analysis, artificial intelligence (AI) and other related theoretical concepts are analyzed and introduced. Then, the decision tree of single classification algorithm and the random forest (RF) of ensemble learning algorithm are analyzed, and the academic performance prediction model of online education is constructed by RF algorithm. Finally, the data of education platform is used for empirical analysis to verify the reliability and practicability of the academic performance prediction algorithm of online education. The connotation of learning analysis, the role and elements of learning analysis in the learning process are introduced. The algorithm principle of RF and decision tree is analyzed. By using the idea of information entropy and discretization, the continuous variables are processed to improve the fitting degree of the algorithm. The model is evaluated by empirical analysis, and the test accuracy of several different algorithms is compared. It is found that the prediction accuracy of the RF algorithm is more than 90%, which shows that the prediction method can help teachers and students to carry out better teaching and learning activities, so as to better improve students' ability to master knowledge. It is hoped that the result can provide some reference for the management of students' learning behavior and the optimization of teachers' teaching strategies in online learning activities.

Keywords—Online education; academic performance prediction; RF; decision tree; artificial intelligence

1 Introduction

The development of big data and AI technology has brought changes to people's life style. The situation of teaching and learning has also changed because of the development of Internet and AI technology. The successful use of Massive Open Online Course (MOOC) also brings new development opportunities for the learning form of online learning. The continuous accumulation of relevant data has gradually formed education big data, which makes the education field face the severe challenge of data-driven mode [1]. Online learning behavior refers to the set of learning related behaviors that occur in the network learning environment. Online education has become an important way of teaching and learning. However, due to the special form of online learning, it is difficult to solve the problems in teaching quality, teaching personalization, teaching monitoring and teaching evaluation [2]. In foreign countries, the research on learning analysis started relatively early, and the scope of theoretical research and empirical research on learning analysis in the field of education is wide. Many scholars use the regression analysis method, and at the same time, they also put forward several analysis frameworks and models. After analyzing the research results in this field, some scholars summarized learning analysis methods into five categories: statistical analysis and visualization, clustering, text mining, relationship mining and prediction [3]. Relatively speaking, the research on data mining and learning analysis started late in China. Most scholars in China have studied the width of learning analysis, and used learning analysis to predict online learning behavior. Through data mining technology and machine learning methods, some scholars have compared and analyzed the prediction effect of single classifiers and integrated classifiers, established the academic performance prediction model of online learning, and proved that the integrated learning algorithm can be used for the construction of the classification model, [4].

Through data mining and analysis of learners' learning characteristics, some scholars built a learner learning characteristics analysis system combined with the real situation, and successfully designed the analysis and evaluation system of online learning behavior characteristics, which can help educators better grasp the learning behavior information and optimize teaching methods [5]. In terms of the prediction of students' academic performance, some scholars use clustering analysis and decision tree classification algorithm to predict college students' English scores under the network teaching mode, and provide students with personalized learning environment through the prediction results [6]. Through the research of data mining algorithms, many scholars have successfully constructed a performance prediction system, which provides effective strategies for the reform of teaching management and the improvement of teaching quality [7].

Based on the existing research results, the related theoretical concepts such as learning analysis and AI are introduced. Then, the decision tree of single classification algorithm and random forest (RF) of ensemble learning algorithm are analyzed, and the prediction model of online education academic performance is constructed by RF algorithm. Finally, an empirical analysis is carried out using the data of the education platform to verify the reliability and practicability of the academic performance prediction algorithm of online education.

2 Method

2.1 AI

AI is to make machines think and behave like humans so that they can complete deeper work with human intelligence [8]. AI is a comprehensive subject which integrates the knowledge of many subjects. It involves computer science, psychology, mathematics, statistics, linguistics and so on. It is assumed to cover the knowledge of all disciplines in natural science and social sciences. AI can be strong and weak. Strong AI refers to the intelligent machine which can make real reasoning and solve problems independently. This kind of intelligent machine has the ability of self-awareness and perception. At present, the AI technology that people study is mainly weak AI. Weak AI means that machines cannot produce human-like thinking, that is to say, people cannot really produce intelligent machines that can independently reason and solve problems. Although this kind of machine seems to have intelligent thinking, in fact, it will not have autonomous consciousness and intelligence similar to human beings. Strong AI and weak AI are not completely opposite. Even if it can be fully realized, strong AI cannot completely replace weak AI [9].

The research fields of AI include knowledge representation, machine learning, artificial neural network, expert system and pattern recognition. The objective of the research is to make the computer have the learning ability similar to human beings [10]. AI decision-making method is to use the idea of knowledge representation and processing in AI to make decisions. At the same time, with the help of the methods of applied management science, computer science and related disciplines, the decision-making schemes are analyzed and compared, so as to help managers make correct decisions.

2.2 Learning analysis

At present, there is no unified definition for learning analysis. Learning analysis mainly focuses on learners' learning process and learning environment. With the help of existing data sets for modeling and analysis, learners' learning rules or learning performance can be obtained, which can be used for feedback and intervention to help them learn more effectively. The data sources of learning analysis mainly include real-time classroom teaching data, online learning platform data and data recorded by education management system. The object of learning analysis is learners' learning. Online learning process mainly includes learning time, learning content, learning results and learning tools. Learning goal is the guidance of learning, learning tools assist learning activities, and the platform records and stores learners' operation behavior. The participation of learning tools and learning activities directly affect the learning results of learners, that is, it is related to students' academic performance. Learning analysis can provide reference for teaching evaluation and teaching scheme, and analyze which learning elements can make learning really effective, thus helping to formulate teaching methods and teaching strategies [11].

The behavior of learners in the learning process can affect their academic performance. The main content is the prediction and analysis of learners' performance in online education. Prediction analysis is carried out mainly through the learner's own ability, research process, research purpose, algorithm and data set, as well as data analysis support.

2.3 RF algorithm

RF algorithm is a homogeneous ensemble learning algorithm. RF is developed on the basis of decision tree. It selects samples and feature attributes through the idea of random selection [12]. RF algorithm can deal with the data with noise and missing values, and has a fast data fitting process. Then, the measurement of the importance of features in RF can provide a basis for feature selection, which has a wide range of applications.

First, the decision tree of base learner for RF is analyzed. The decision tree has a tree structure, which classifies data according to a series of rules. The decision tree generally includes three types of nodes: root node, internal node and leaf node13]. The general process of building decision tree is to select the root node first, that is to determine the initial split attribute. The training set generally contains the correct classification of the record set, which contains the information required by machine learning. The decision tree algorithm takes the best attribute in the training set as the root node, and then finds each sub node through recursion. Then, on the basis of finding the best attribute, other nodes are continued to be found by selecting the root node until a complete tree is formed. At present, there are three kinds of decision tree construction algorithms: ID3, C4.5 and CART. The first two algorithms are the core of decision classification algorithms [14]. The construction methods are as follows. Firstly, the optimal partition feature is found and regarded as the node, and then the dataset is divided into several parts according to each value of the feature. Then, the subset is divided by recursive algorithm. Finally, whether it meets the termination condition of recursion is tested.

Information entropy can be regarded as the system transforming the information source data into state representation. Information source data is a random event with uncertainty, which is measured by probability P. If information source U can have n values (U1, U2, …, Un), the corresponding probability is P1, P2, …, Pn, and the events are independent of each other. The average uncertainty of information source is described as information entropy, which is expressed by the equation 1.

$$I(U) = -\sum_{i=1}^{n} P_i log_2 P_i \tag{1}$$

Pi in the above equation represents the i-th probability value. Information entropy can be used to measure the uncertainty of a category, as well as the uncertainty of a feature. The larger the information entropy of eigenvector is, the more chaotic it is.

The information gain is the difference between the two information entropies, which represents the change of the chaos degree between the nodes in the decision tree. When the data set is S and the total number of samples is s, the calculation method of the original information entropy is as equation 2.

$$I(S_1, S_2, \dots, S_m) = -\sum_{i=1}^{m} P_i log_2 P_i \tag{2}$$

m in the above equation refers to the number of categories, and Sm represents the number of samples belonging to a certain category. In order to find the optimal feature of the whole data set, the information entropy of each feature in the data set needs to be calculated. The empirical information entropy of attribute A is calculated as equation 3.

$$E(A) = \sum_{j=1}^{v} \frac{S_{1j}+S_{2j}+\cdots S_{mj}}{S} \mathrm{I}(S_{1j} + S_{2j} + \cdots S_{mj}) \tag{3}$$

The information entropy of feature A is calculated on the eigenvalue, that is, the subset is divided according to the eigenvalue. In order to express the uncertainty of features by information entropy, the information entropy I (S1j, S2j, ... Smj) of each subset needs to be calculated, as shown in equation 4.

$$\mathrm{I}\left(S_{1j} + S_{2j} + \cdots S_{mj}\right) = -\sum_{i=1}^{m} P_{ij} log_2 P_{ij} \tag{4}$$

In the above equation, v refers to the number of eigenvalues, Smj represents the number of samples of m class under the j-th eigenvalue, (S1j + S2j +... Smj) represents the weight of j feature. Therefore, the calculation expression of information gain is as shown in equation 5.

$$\mathrm{Gain(A)} = \mathrm{I}\left(S_{1j} + S_{2j} + \cdots S_{mj}\right) - \mathrm{E(A)} \tag{5}$$

The final decision tree has a good classification effect on the data in the training set, but the prediction accuracy of the unknown data set needs to be further improved. In the classification of test sets, the complexity of decision tree model will lead to over fitting, and finally lead to the inaccuracy of classification.

When the RF contains N base classifiers (decision trees), each decision tree contains one classification result, so for the same test set, N classification results will be generated. RF mainly introduces random thinking into the process of model construction [15]. The method of building RF model is as follows. If the initial data set contains m samples, m samples are obtained by m times of sampling with return. In this process, due to the randomness, there is a certain difference in the number of samples collected. The m samples collected are called sampling set. By repeating this process t times, T sample sets with m samples can be obtained. Then, the sample features are randomly selected. If the number of features in the current feature set is V, different from the decision tree, the RF algorithm randomly selects K feature sets from V feature sets, and selects the optimal one as the splitting point. The value of K is random. When its value is equal to V, the construction process is consistent with that of decision tree. When its value is 1, a feature can be randomly selected as the splitting point.

RF improves the diversity of decision tree by using random idea, and enhances the final generalization ability of the algorithm. With the continuous increase of integration scale, RF generally converges to relatively low generalization error, as stated in [16].

2.4 Construction of academic performance prediction model of online education based on RF algorithm

This research is mainly completed with the help of Scikit-leam (sklearn) toolkit based on Python language [17]. The toolkit contains a variety of commonly used machine learning methods, and regression, clustering and preprocessing modules. The toolkit runs in the interactive notebook environment of Jupyter Notebook, which can be used for real-time code writing, and facilitate information sharing and visualization [18]. The construction process of online learning academic performance prediction model is shown in Figure 1.

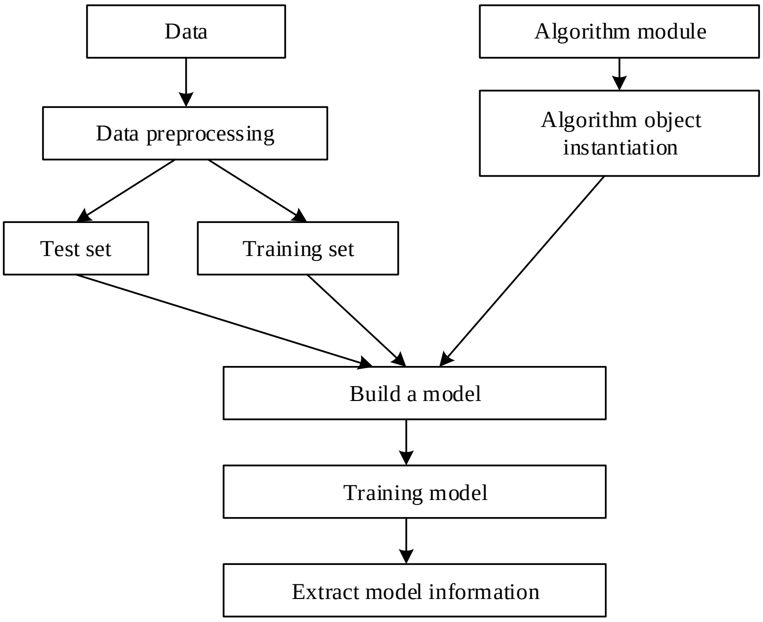

Fig. 1. The flow chart of the construction of online academic performance prediction model

2.5 Data collection and processing

The research data is mainly from some learners' learning data in the database of an online course learning platform. The data include the data records of users' learning on the platform, users' basic information, course selection information, participation in discussions, academic performance and many other information data. The original data contain a lot of information. However, there will also be useless and redundant information in these data. Therefore, it is necessary to preprocess these data, remove the missing data from the user behavior data, delete and filter the behavior data with empty user name, time and other fields. Data preprocessing means mainly include data cleaning, feature selection and data conversion. The purpose of data processing is to improve data quality and make data better adapt to specific data mining tools and methods [19].

Sklearn tool is used to build the intelligent model to deal with numerical data. Therefore, it is also necessary to convert the features of the data set into numerical modes, which can be directly encoded, and can also be encoded by the way of independent hot coding. Then, it is necessary to discretize the continuous features of the obtained data. Discretization is to divide a continuous interval into several discrete spaces. After the data is discretized, its feature representation is closer to the knowledge level, which makes the data easier to be understood and interpreted. The process of data discretization is generally as follows. The preprocessing data is sorted, and then the segmentation breakpoint is determined to determine whether it is necessary to continue dividing the space [20].

3 Results and Analysis

3.1 Determination of feature attributes in algorithm testing

The data features of online learning platform are analyzed, and their weights are analyzed by RF algorithm. Through analysis, a total of 12 feature attributes are obtained, and the weights of these 12 feature attributes are shown in Figure 2.

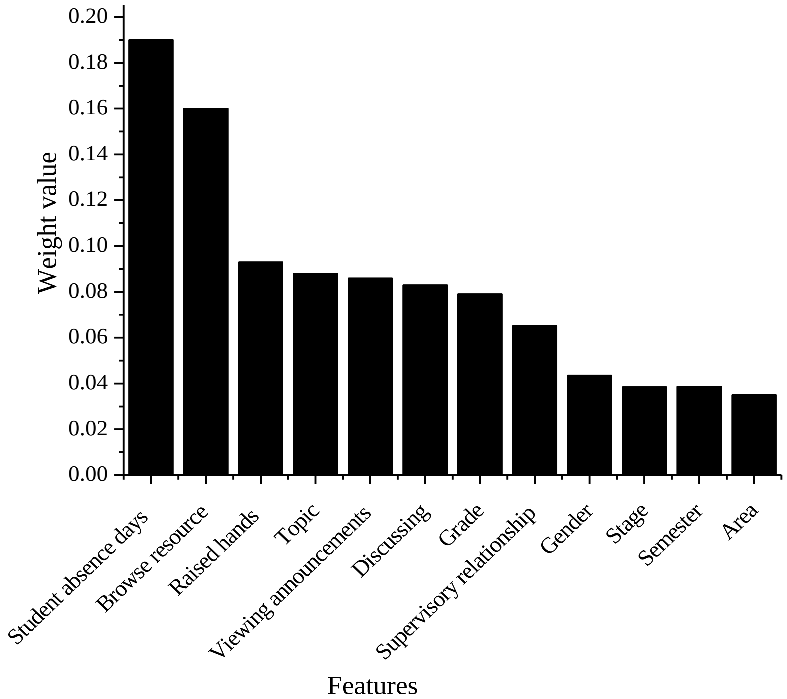

Fig. 2. Weight analysis of each feature attribute

It can be seen that the weight value of the last four feature attributes is very small, which indicates that these features are not important, and can be deleted to reduce the calculation amount of the algorithm. The test results after deletion are compared with those before, and the comparison results are shown in Figure 3.

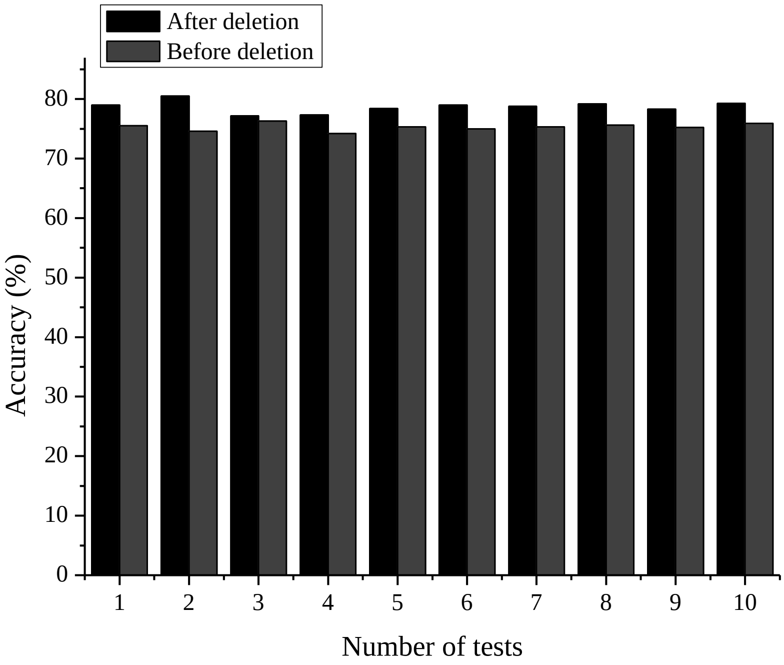

Fig. 3. Comparison of test results before and after eigenvalue deletion

It can be seen that the test accuracy of the algorithm is improved after the eigenvalues which have little impact on the prediction and classification results of the algorithm are deleted, which shows that deleting these features can reduce the calculation amount of the algorithm and optimize the classification and prediction effect of the algorithm.

3.2 Performance test of the algorithm

The number of parameters in the RF is also the number of decision trees. Different number of parameters will lead to different accuracy of the algorithm. Generally speaking, in a certain range, the larger the parameter value, the better the prediction accuracy of the RF algorithm. Within a certain range, the relationship between the accuracy of the RF algorithm and the number of parameters is shown in Figure 4.

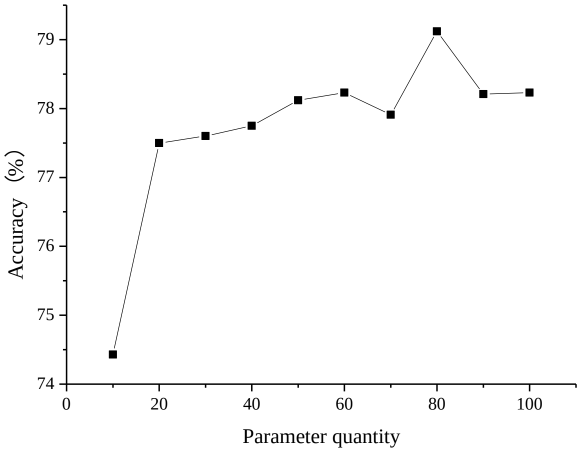

Fig. 4. The curve of the relationship between the accuracy of RF algorithm and parameter quantity

It can be seen that with the increase of parameters, the test accuracy of RF algorithm is also improved. When the parameter quantity is 80, the algorithm has the highest accuracy. When the number of parameters is further increased, the accuracy of the algorithm decreases slightly and does not increase any more.

In order to verify the prediction performance of the RF model, the prediction effect of the RF model is compared with the decision tree and the commonly used support vector machine classification method. The comparison results are shown in Figure 5.

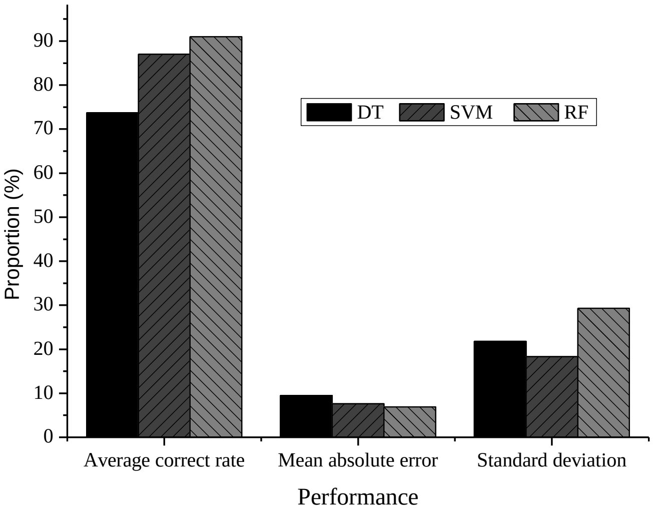

Fig. 5. Comparison and analysis of the performance test of different algorithms

It can be seen that the prediction accuracy of single classification decision tree is the lowest, the prediction accuracy of integrated classification RF algorithm is the highest, and the prediction effect of support vector machine is close to that of RF. The average absolute error of RF algorithm is also the smallest, which shows that the RF algorithm can improve the prediction effect of the algorithm and improve the recognition accuracy of learning behavior analysis.

3.3 Analysis of the relationship between learning behavior characteristics and students' academic performance

The students' academic performance is divided into four grades: excellent (90-100), good (80-89), medium (60-79) and poor (0-59), with 100 students in each grade. The influence of the characteristics of students' participation (days of absence, times of raising hands, number of discussions, times of browsing resources and times of viewing announcements) on students' academic performance is analyzed. The results are shown in Figure 6 and Figure 7.

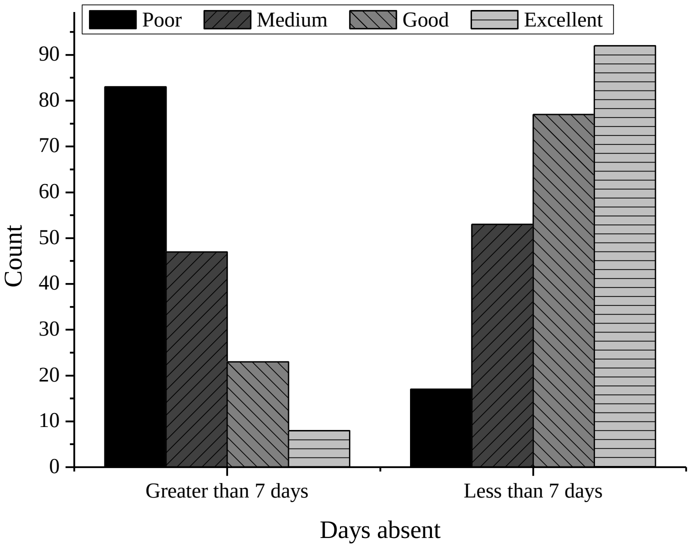

Fig. 6. The influence of absence days on academic performance

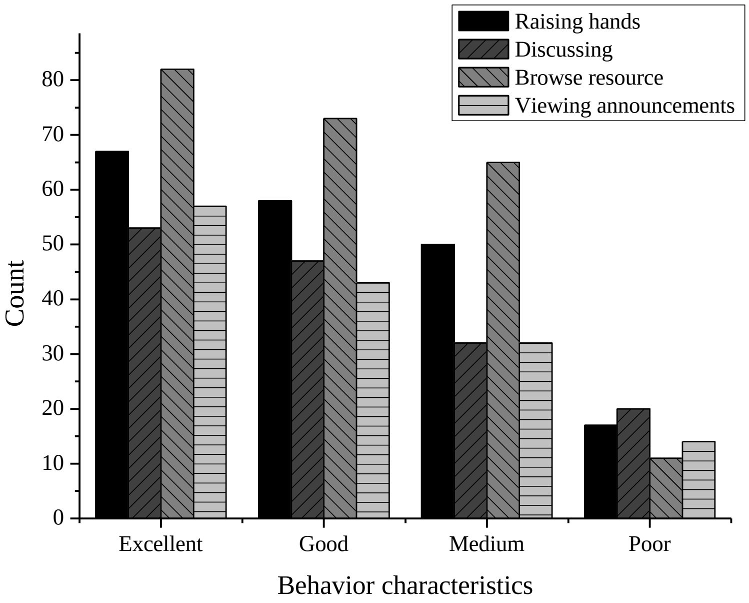

Fig. 7. The relationship between the characteristics of students' participation in learning and their academic performance

It can be seen that there is an inverse relationship between the number of days absent and academic performance. Absences of more than 7 days are more common among students with poor performance. The better a student's grades are, the fewer days a student will be absent. It can be seen that the number of raising hands, the number of discussions, the number of browsing resources and the number of viewing announcements will affect students' academic performance. These factors are directly proportional to students' academic performance. The more often a student exhibits these behavioral characteristics, the better the student performs.

4 Conclusion

The prediction methods of students' online academic performance are mainly studied. First, learning analysis, AI and other related theoretical concepts are analyzed and introduced, the connotation of learning analysis and the role of learning analysis in the learning process, components and the impact on students' academic performance are introduced. Then, the decision tree of single classification algorithm and the RF of ensemble learning algorithm are analyzed, and the academic performance prediction model of online education is constructed by RF algorithm. The algorithm principle of RF and decision tree is analyzed. In order to improve the fitting degree of the algorithm, continuous variables are processed by using the algorithm of information entropy and discretization. Finally, the data of the education platform is used for empirical analysis to verify the reliability and practicability of the academic performance prediction algorithm of online education. The RF algorithm is used to predict the students' performance. At the same time, the relationship between the characteristics of learning behavior and students' academic performance is analyzed.

Due to limited knowledge, the writing is not profound, and the model selection is not deep enough. At the same time, the analysis on the relationship between students' learning characteristics is insufficient, so it cannot comprehensively demonstrate the learning state of online learners. It is hoped that in the follow-up research, the prediction algorithm can be deeply analyzed and selected to optimize the academic performance prediction method of online education.

5 References

[1] Mingsiritham, K. Chanyawudhiwan, G. (2020). Experiment of the prototype of online learning resources on massive open online course (mooc) to develop life skills in using technology media for hearing impaired students. International Journal of Emerging Technologies in Learning (iJET): 15(3): 242. https://doi.org/10.3991/ijet.v15i03.12059

[2] Prior, D. D. Mazanov, J. Meacheam, D. Heaslip, G. Hanson, J. (2016). Attitude, digital literacy and self-efficacy: flow-on effects for online learning behavior. Internet & Higher Education, 29(apr.): 91-97. https://doi.org/10.1016/j.iheduc.2016.01.001

[3] Brohn, & D., M. (1981). The use of the plane frames program as an aid to learning in structural analysis. Computers & Education, 5(1): 37-44. https://doi.org/10.1016/0360-1315(81)90025-7

[4] Yang, F. Li, F. W. B. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. Computers & Education, 123(AUG.): 97-108. https://doi.org/10.1016/j.compedu.2018.04.006

[5] Liu, S. Ni, C. Liu, Z. Peng, X. Cheng, H. N. H. (2017). Mining individual learning topics in course reviews based on author topic model. International Journal of Distance Education Technologies, 15(3): 1-14. https://doi.org/10.4018/ijdet.2017070101

[6] Bharara, S. Sabitha, S. Bansal, A. (2018). Application of learning analytics using clustering data mining for students' disposition analysis. Education and Information Technologies, 23(2): 957-984. https://doi.org/10.1007/s10639-017-9645-7

[7] Cohen, A. Holstein, S. (2018). Analysing successful massive open online courses using the community of inquiry model as perceived by students. Journal of Computer Assisted Learning, 34(5): 544-556. https://doi.org/10.1111/jcal.12259

[8] Lawless, W. F. Mittu, R. Sofge, D. Hiatt, L. (2019). Artificial intelligence, autonomy, and human-machine teams — interdependence, context, and explainable ai. Ai Magazine, 40(3): 5-13. https://doi.org/10.1609/aimag.v40i3.2866

[9] Hameed, M. Sharqi, S. S. Yaseen, Z. M. Afan, H. A. Hussain, A. Elshafie, A. (2017). Application of artificial intelligence (ai) techniques in water quality index prediction: a case study in tropical region, malaysia. Neural Computing and Applications, 28(1): 893-905. https://doi.org/10.1007/s00521-016-2404-7

[10] Kulkarni, R. H. Padmanabham, P. (2017). Integration of artificial intelligence activities in software development processes and measuring effectiveness of integration. Iet Software, 11(1): 18-26. https://doi.org/10.1049/iet-sen.2016.0095

[11] Kim, K. Moon, N. (2019). Activity index model for self-regulated learning with learning analysis in a tel environment. Journal of supercomputing, 75(4): 1971-1989. https://doi.org/10.1007/s11227-018-2446-y

[12] Bai, S. (2017). Growing random forest on deep convolutional neural networks for scene categorization. Expert Systems with Applications, 71(APR.): 279-287. https://doi.org/10.1016/j.eswa.2016.10.038

[13] Krishnakumari, A. Elayaperumal, A. Saravanan, M. Arvindan, C. (2017). Fault diagnostics of spur gear using decision tree and fuzzy classifier. International Journal of Advanced Manufacturing Technology, 89(9-12): 1-8. https://doi.org/10.1007/s00170-016-9307-8

[14] Zhang, K. Wu, X. Niu, R. Yang, K. Zhao, L. (2017). The assessment of landslide susceptibility mapping using random forest and decision tree methods in the three gorges reservoir area, china. Environmental Earth ences, 76(10): 405. https://doi.org/10.1007/s12665-017-6731-5

[15] Ana, D. C. Torres-Sánchez Jorge, Jose, P. Jiménez-Brenes Francisco, Ovidiu, C. López-Granados Francisca. (2018). An automatic random forest-obia algorithm for early weed mapping between and within crop rows using uav imagery. Remote Sensing, 10(3): 285-. https://doi.org/10.3390/rs10020285

[16] Thanh Noi, P. Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. Sensors, 18(1): 18. https://doi.org/10.3390/s18010018

[17] Mcginnis, W. D. Siu, C. Andre, S. Huang, H. (2018). Category encoders: a scikit-learn-contrib package of transformers for encoding categorical data. The Journal of Open-Source Software, 3(21): 501. https://doi.org/10.21105/joss.00501

[18] Suarez, A. Alvarez-Feijoo, M. A. Fernandez Gonzalez, R. Arce, E. (2018). Teaching optimization of manufacturing problems via code components of a jupyter notebook. Computer Applications in Engineering Education, 26(5): 1102-1110. https://doi.org/10.1002/cae.21941

[19] Ouyang, Q. Lu, W. (2018). Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods. Water Resources Management, 32(2): 1-16. https://doi.org/10.1007/s11269-017-1832-1

[20] Liu, X. Marchis, L. Debiase, E. Breaux, K. C. Courville, T. Pan, X. et al. (2016). Do cognitive patterns of strengths and weaknesses differentially predict errors on reading, writing, and spelling? Journal of Psychoeducational Assessment, 35(1-2): 186-205. https://doi.org/10.1177/0734282916668996

6 Author

Jing Yu is a lecturer and master of Zhejiang Business College, Mainly engaged in student behavior management and online education research.

Article submitted 2020-12-03. Resubmitted 2021-01-11. Final acceptance 2021-01-12. Final version published as submitted by the authors.