IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 April 2023, accepted 16 May 2023, date of publication 22 May 2023, date of current version 30 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3278370

RESEARCH ARTICLE

Analysis of Learning Behavior Characteristics and Prediction of Learning Effect for Improving College Students' Information Literacy Based on Machine Learning

YONG SHI[1], FANG SUN[2], HONGKUN ZUO[3], AND FEI PENG[1]

[1]School of Computer Science, Huainan Normal University, Huainan 232038, China
[2]Institute of Education, Huainan Normal University, Huainan 232038, China
[3]School of Finance and Mathematics, Huainan Normal University, Huainan 232038, China

Corresponding author: Yong Shi (shiyong@hnnu.edu.cn)

This work was supported in part by the Scientific Research Project of China Anhui Provincial Department of Education under Grant 2020jyxm1733 and Grant 2020qkl38, in part by the Key Research Project of Huainan Normal University under Grant 2022XJZD027.

ABSTRACT Information literacy is a basic ability for college students to adapt to social needs at present, and it is also a necessary quality for self-learning and lifelong learning. It is an effective way to reveal the information literacy teaching mechanism to use the rich and diverse information literacy learning behavior characteristics to carry out the learning effect prediction analysis. This paper analyzes the characteristics of college students' learning behaviors and explores the predictive learning effect by constructing a predictive model of learning effect based on information literacy learning behavior characteristics. The experiment used 320 college students' information literacy learning data from Chinese university. Pearson algorithm is used to analyze the learning behavior characteristics of college students' information literacy, revealing that there is a significant correlation between the characteristics of information thinking and learning effect. The supervised classification algorithms such as Decision Tree, KNN, Naive Bayes, Neural Net and Random Forest are used to classify and predict the learning effect of college students' information literacy. It is determined that the Random Forest prediction model has the best performance in the classification prediction of learning effect. The value of Accuracy is 92.50%, Precision is 84.56%, Recall is 94.81%, F1-Score is 89.39%, and Kapaa coefficient is 0.859. This paper puts forward differentiated intervention suggestions and management decision-making reference in the information literacy teaching process of college students, with a view to adjusting the information literacy teaching behavior, improving the information literacy teaching quality, optimizing educational decision-making, and promoting the sustainable development of high-quality and innovative talents in the information society.Our work involving research of the thinking and direction of the sustainable development of information literacy training proved to be encouraging.

INDEX TERMS Machine learning, information literacy, learning behavior characteristics, learning effect, innovative talents.

I. INTRODUCTION

With the rapid development of information technology represented by computer, network technology and communication technology, computers and the Internet have been widely used in various fields of society. Information plays an increasingly important role in the development of human society and increasingly becomes one of the most active and decisive factors in all fields of society. Information literacy, critical thinking and creativity are the core skills that college students

The associate editor coordinating the review of this manuscript and approving it for publication was James Harland.

must master in the 21st century [1]. In the information age, information literacy is an important part of college students' core literacy. Information literacy is a kind of adaptability to the information society. The information literacy of college students is directly related to the sustainable development of future talents and the cultivation of innovative talents [2], [3].Information literacy is a part of cultural literacy and overall quality. Cultivating college students' information literacy has already become an important issue facing contemporary higher education.

Information literacy includes the basic knowledge and skills of information and information technology, the ability to use information technology to learn, cooperate, communicate and solve problems, as well as information awareness and social ethics. At present, information literacy education has received the attention of people from all walks of life. The education departments and libraries in the United States, the United Kingdom, Australia and other countries have carried out information literacy education to different degrees.In 2022, the Ministry of Education and other four departments of China jointly issued the ''key points of improving the digital literacy and skills of the whole people in 2022''. Students' information literacy and digital literacy are expected to be further improved in the next few years [4].In recent years, due to the influence of online teaching and hybrid teaching, and the development of artificial intelligence technology, information literacy has also received more and more research attention. Many colleges and universities at home and abroad have opened information literacy courses through various ways to carry out targeted information literacy education. For example, on the MOOC platform of the University of China, Tsinghua University has opened ''Information Literacy: A Compulsory Course for Academic Research'', Wuhan University has opened ''Information Literacy and Practice - A Pair of Academic Eyes'', Sun Yat-sen University's ''Information Literacy General Course - A Compulsory Course for Digital Survival'', and Sichuan Normal University's ''Information Literacy and Lifelong Learning (Autonomous Mode)'' [5].In view of the existing information literacy education for college students, many problems have emerged.

In the field of education big data, learning prediction is a very meaningful topic. Learning effect prediction is one of the core issues in the field of learning analysis. Its essence is to use various data generated by learners in the learning process, and use the method represented by machine learning to predict the learning effect. According to the prediction results, teachers can know the learners' learning status in time and intervene in the learning process in time. Such as improving learners' learning habits, adjusting teaching strategies, etc. Wufati and Hao [6].Learning analysis technology has developed from principle exploration and application value to application in learning behavior analysis, data visualization and learning prediction Hang et al. [7]. Learning prediction is based on learning achievement, learning goals, and learning ability, and predicts learning effect and learning experience based on the characteristics of learning behavior before and after learning AlShammari et al. [8].The prediction of learning results includes prediction theoretical model, empirical research of prediction model, comparison of algorithms, development of algorithms, research of early warning factors and literature review, etc. The prediction of students' learning performance and learning effect is carried out using regression analysis, neural network, Bayes and other methods Gaihua and Gangshan [9]. UNESCO's 2019 report, Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development, explores how artificial intelligence technologies can help education systems use data to to promote equity and quality in education [10].Using educational data mining technology and machine learning technology to build learning effect prediction model through data-driven way, that is, automatically learning from data to build prediction model, which is the current research focus and research trend.

This study links multiple specific behavioral data together to create an integrated data link based on college students' learning behaviors in information literacy courses. The predictive analysis and evaluation of different machine learning classification models are used to classify and predict the learning effect of college students. This study focuses on the following questions.

(1) Which indicators of information literacy learning behavioral characteristics of college students have better predictive ability for learning effect?

(2) Which machine learning models have better predictive performance and efficacy based on the study sample?

(3) What diagnostic observations for use in learning recommendations and instructional interventions were derived in conjunction with the study findings?

II. LITERATURE REVIEW

This study conducted a literature study on the analysis of learning behavior characteristics and prediction of learning effect to improve college students' information literacy.

The literature data mainly comes from the common databases for international paper retrieval such as Web of Science, Scopus, Ei Compendex, etc., and is mainly based on the relevant research in the past three years.

Since there are many professional terms and machine learning algorithm terms in the reference documents, they are uniformly described, as shown in Table 1.

A. INFORMATION LITERACY LEARNING BEHAVIOR ANALYSIS AND LEARNING EFFECT EVALUATION

Scholars have carried out research on information literacy from different angles. Specific literature analysis and comparison are as follows:

Literature comparison in research domain: In terms of information literacy learning behavior and learning effect, many studies in recent years have focused on the evaluation framework of information literacy effect; Strategies for improving information literacy learning behavior; The

TABLE 1. Abbreviation of professional terms.

Acronyms	Description
M L	Machine Learning
ITUB	Information Technology Usage Behaviour
PSCLE	Preferences for Smart Classroom Learning Environments
ILSs	Information Literacy Skills
ISB	Information-Seeking Behavior
CBL	Case-Based Learning
DNN	Deep Neural Network
CNN	Convolutional Neural Network
SVM	Support Vector Machine
GBDT	Gradient Boosting Decision Tree
LR	Logistic Regression
RF	Random Forest
KNN	K-Nearest Neighbor
BP-NN	Back Propagation Neural Network

cultivation of specific ability of information literacy; Online courses or the relationship between intelligent environment and information literacy.

Literature comparison on research methodology: Most scholars mainly use quantitative research, qualitative research, questionnaire survey, data mining, and factory quality-experience, while some scholars use machine learning technology. The tick marks in Table 2 represent the use of machine learning. According to the literature, scholars mainly use traditional research methods in research methods, and seldom use machine learning methods to carry out relevant research.

Literature comparison in the research of finding: Scholars have made fruitful exploration results in the research. For example, some scholars establish a correct teaching mode, some scholars explore new teaching methods of information literacy, and some scholars build prediction models to enhance the use of information literacy. The literature shows that there are few achievements in the analysis of information literacy learning behavior and the construction of learning effect prediction model.The detailed comparative studies are summarized in Table 2.

B. LEARNING BEHAVIOR ANALYSIS AND LEARNING EFFECT PREDICTION BASED ON MACHINE LEARNING

Scholars have carried out research on learning behavior analysis and learning effect from different angles. Specific literature analysis and comparison are as follows:

Literature comparison in research domain: The research mainly focuses on teaching effect prediction model research, learning performance model research, teaching model effect research, learning quality analysis and curriculum evaluation research.

Comparison of literatures on research methodology: Most research methods are mainly machine learning algorithms, including neural networks, decision trees, support vector machines and other algorithms.The tick marks in Table 3 represent the use of machine learning. Comprehensive analysis and quantitative research were used in individual studies.

Comparison of literature on finding: There are many research results based on machine learning algorithm. For example, performance prediction model; Students' willingness analysis model; Prediction of classroom teaching effect; Learning behavior diagnosis model, etc.

Scholars have carried out research on learning behavior analysis and learning effect from different angles. Specific literature analysis and comparison are as follows:

Literature comparison in research domain: The research mainly focuses on teaching effect prediction model research, learning performance model research, teaching model effect research, learning quality analysis and curriculum evaluation research.

Comparison of literatures on research methodology: Most research methods are mainly machine learning algorithms, including neural networks, decision trees, support vector machines and other algorithms; Comprehensive analysis and quantitative research were used in individual studies.

Comparison of literature on finding: There are many research results based on machine learning algorithm. For example, performance prediction model; Students' willingness analysis model; Prediction of classroom teaching effect; Learning behavior diagnosis model, etc.

To sum up, the current research on information literacy is mainly based on theoretical deduction and experience, establishing the hypothesis that some factors are related to academic performance, and then collecting data through questionnaires and interviews to analyze and verify the hypothesis. This method can only prove the correlation between selected factors and academic achievement, but it is difficult to determine the quantitative relationship between selected factors and academic achievement. Machine learning and data mining technology are rarely used, and data intelligence analysis research of information literacy education is lacking. Some researchers use decision tree, neural network and other algorithms to establish academic achievement prediction models, but lack of information literacy learning effect prediction research.

With the continuous development and maturity of intelligent technologies such as data mining, emotion analysis and pattern recognition, especially the combination of machine learning technology and education field, it provides strong technical support for learning prediction research. Although some studies have pointed out the negative impact of artificial intelligence on educational research, the use of educational data mining and other technologies is still the current research trend. Therefore, it is an urgent problem to build an information literacy learning behavior characteristic analysis and

TABLE 2. Research on information literacy learning behavior and learning effect.

Ref.	Domain	M L	Methodology	Finding
[11]	Hot Spots and Enlightenment		Bibliometric analysis	Assessment tools, information security, and personalized learning recommendations
[12]	Evaluation of Enhancement	√	ML	Constructs a predictive model for enhancing ITUB
[13]	Smart Classroom Preferences		A quantitative method	A high level of information literacy obtained significantly higher scores on PSCLE
[14]	Training Strategies		Data mining	Discusses several strategies of information literacy education under the background of big data
[15]	Online Course Based on MOOC		Analyzes the data	The "MOOC+SPOC+Flipped Classroom" teaching method
[16]	Assessment of ILSs and ISB		Quantitative research	Enhancing the ILSs of medical students
[17]	The Frame of Evaluation Index		Means of questionnaires	The methods to improve information literacy
[18]	Assess the Self-ssessment		Statistical free software R	Explore new teaching methods
[19]	Metacognitive Abilities		Factorial quasi-experiment	Improve metacognitive abilities classified based on students' information literacy
[20]	Evaluation of Information Literacy		Quantitative evaluation and data mining	Establish a reciprocal teaching mode
[21]	Critical Thinking Skills		Directed qualitative content analysis	The CBL unit was effective in increasing their information literacy and critical thinking skills
[22]	Cross-Media Data Analysis	√	DNN	Analyzes their media selection tendency, media usage time, positive influence, and the relationship with new media literacy
[23]	A Multilevel Modeling Approach		Multilevel modeling	The models related to teacher and student characteristics

learning effect prediction model for college students with strong usability, easy operation and good prediction performance, as well as differential recommendation and intervention based on the prediction results.

III. MATERIALS AND METHODS

A. RESEARCH TOOLS

Common learning prediction tools include Weka, SPSS, Python, Rapidminer and other tools. In this study, SPSS and Rapidminer analysis tools are mainly used in data preprocessing, feature set selection and classification prediction, and model performance evaluation. Rapidminer is mainly used in machine learning. Rapidminer is the world's mainstream data mining and machine learning software. It provides functions such as data preprocessing and visualization, predictive analysis and statistical modeling, evaluation and deployment, and has rich machine learning algorithms [39].

B. RESEARCH OBJECT AND DATA SOURCE

Due to regional factors, different regions have different requirements for information literacy. Therefore, the establishment of information literacy standards should not be limited to general standards [40].The research team has built the evaluation index system of college students' information literacy in the previous research [41].The index system provides a basic reference tool for this paper.Based on this evaluation index, the research team observed, measured, extracted and described the information literacy learning behavior characteristics of college students, and formed the information literacy learning behavior characteristics observation scale for college students. The scale includes awareness and attitude, knowledge and skills, application and innovation, ethics and responsibility. Awareness and attitude mainly focus on the understanding of the importance of information technology. Knowledge and skills mainly focus on the knowledge and skills of information technology. Application and innovation mainly examine the cognitive thinking and innovative application of information technology. Morality and responsibility mainly focus on information laws, regulations and moral concepts. There are 4 first-level indicators, 9 second-level indicators and 28 third-level indicators. In order to measure the learning effect of students,

TABLE 3. Learning behavior analysis and learning effect prediction based on machine learning.

Ref.	Domain	ML	Methodology	Finding
[24]	Predicting and Analyzing Performance	√	LR, RF, CNN	Several prediction models have been created to predict performance
[25]	Management of Learning Quality of Online Courses	√	A lightweight CNN model	Corners can be used to detect student attention
[26]	Information Anxiety	√	SVM Optimization Alogrithm	The redundant information is filtered through optimization algorithms model
[27]	Prediction Model for the Teaching Effect	√	Apriori algorithm	Predicting the effect of Two courses classroom teaching
[28]	Prediction Model Innovation and Entrepreneurship	√	GBDT,LR	A model for analyzing students' willingness
[29]	Enhanced learning	√	Several ML algorithms	Numeracy and Literacy Aptitude Analysis and Prediction
[30]	Behavior Analysis and Management	√	K-means	Data are selected to describe the student's behavior
[31]	Effect of College English Blended Teaching Mode		Comprehensive analysis	Aimed to investigate the students' learning effect
[32]	Perception of Electronic Learning		Quantitative model	Indicated e-learning perceptions' in knowledge mastery, social competence, and media literacy abilities
[33]	Analysis of Practical Teaching Effect	√	Association analysis algorithm	Assist the training teachers to strengthen management
[34]	Technology Integration in Teaching-Learning Practices		Systematic Review	Technology-incorporated teaching effectively enhances teaching practice
[35]	Course Assessment	√	Several ML algorithms	Assess the effect of engagement on student performance
[36]	Hybrid educational data mining model	√	Data mining and ML	Proposed model evaluates the student performances based on distinctive factors
[37]	Learning Performance Prediction	√	Five ML models	A learning behavior diagnosis model combining decision tree and DNN
[38]	Prediction Index of Learning Results	√	BP-NN ,decision tree,Naive Bayes	Timely and effective teaching intervention

this study divides the students' learning scores into five categories: excellent (5), good (4), medium (3), qualified (2) and unqualified (1). Each three-level indicator of information literacy learning behavior characteristics of college students corresponds to Likert's five-level scale: "1=never"; "2=seldom"; "3=sometimes"; "4=often"; "5=always".

Table 4 describes the observed indicators of information literacy learning behavioral characteristics of college students. In conjunction with the above, the four areas of learning behavior are described in terms of learning behavior in consciousness and attitude, learning behavior in knowledge and skills, learning behavior in application and innovation, and learning behavior in morality and responsibility.

Learning behavior in consciousness and attitude:Mainly including Information perception consciousness (IPC), Information application consciousness(IAC) and Lifelong learning consciousness (LLC). Specific behaviors include: Identify and classify information (IPC1); Using the Web to find, filter, and judge information (IPC2); Determine the correctness and reliability of information sources (IPC3); Using information technology related knowledge and methods to solve problems (IAC1); Using information technology tools such as mind mapping tools to assist learning (IAC2); Leveraging Information Technology to support Lifelong Learning (LLC1); Using Information Technology to Support Professional and Personal Development (LLC2).

Learning behavior in knowledge and skills:Mainly including Information science knowledge(ISK) and Information application Skills (IAS). Specific behaviors include: Understand all kinds of operating systems, word processing software, graphics and image processing software, video and audio processing software operation method (ISK1); Understand the development history, basic status and future trend of information technology (ISK2); Master the basic knowledge

TABLE 4. Observation scale of information literacy learning behavior characteristics of college students.

First level indicators	Second level indicators		Observable behavior
Consciousness and Attitude	1	IPC	1) Identify and classify information (IPC1)
			2) Using the Web to find, filter, and judge information (IPC2)
			3) Determine the correctness and reliability of information sources (IPC3)
	2	IAC	4) Using information technology related knowledge and methods to solve problems (IAC1)
			5) Using information technology tools such as mind mapping tools to assist learning (IAC2)
	3	LLC	6) Leveraging Information Technology to support Lifelong Learning (LLC1)
			7) Using Information Technology to Support Professional and Personal Development (LLC2)
Knowledge and Skills	1	ISK	8) Understand all kinds of operating systems, word processing software, graphics and image processing software, video and audio processing software operation method (ISK1)
			9) Understand the history, basic status and future trend of information technology (ISK2)
			10) Master the basic knowledge and technology of information retrieval and evaluation, information classification and storage method (ISK3)
			11) Master the basic scientific knowledge of information literacy, data literacy, visual literacy and other multi-literacy (ISK4)
	2	IAS	12) Use various search engines and network platforms to find the required information (IAS1)
			13) Classify the information and present the information in a tabular form (IAS2)
			14) Identification and analysis of information through various approaches and methods (IAS3)
			15) Create valuable information resources based on specific teaching content or topics (IAS4)
Application and Innovation	1	IT	16) Define and identify implicit assumptions in information, and deduce information (IT1)
			17) Carry out targeted information-based instructional design and implement effective instructional activities (IT2)
			18) Using information technology to support services and management (IT3)
			19) Construct problem solutions by integrating resources and using reasonable algorithms (IT4)
	2	IB	20) Use collaborative tools to create and manage content (such as project management systems, shared documents, etc.) (IB1)
			21) Use advanced communication tools to communicate with people (e.g. video conferencing, data sharing, application sharing) (IB2)
			22) Developing innovative teaching applications (IB3)
			23) To carry out information technology cooperation and exchange (IB4)
Morality and Responsibility	1	IE	24) Healthy and correct use of learning resources to create information environment (IE1)
			25) Restrain one's own information ethical behavior and supervise others' information behavior (IE2)
			26) Abide by the network civilization convention, purify the network language, civilized and polite learning and communication (IE3)
	2	ILR	27) Impart knowledge of laws, regulations and ethics related to technology utilization (ILR1)
			28) Learn the right to access and access information equally and respect the intellectual property rights of others (ILR2)

and technology of information retrieval and evaluation, information classification and storage method (ISK3); Master the basic scientific knowledge of information literacy, data literacy, visual literacy and other multi-literacy (ISK4); Use various search engines and network platforms to find the required information (IAS1); Classify the information and present the information in a tabular form (IAS2); Identification and analysis of information through various approaches and methods (IAS3); Create valuable information resources based on specific teaching content or around specific teaching topics (IAS4).

Learning behavior in application and innovation:Mainly including Information thinking (IT) and Information behavior (IB). Specific behaviors include: Define and identify implicit assumptions in information, and deduce information (IT1); Carry out targeted information-based instructional design and implement effective instructional activities (IT2); Using information technology to support services and management (IT3); Construct problem solutions by integrating resources and using reasonable algorithms (IT4); Use collaborative tools to create and manage content,such as project management systems, shared documents, etc. (IB1); Use advanced communication tools to communicate with people (e.g. video conferencing, data sharing, application sharing) (IB2); Developing innovative teaching applications (IB3); To carry out information technology cooperation and exchange (IB4).

Learning behavior in morality and responsibility: Mainly including Information ethics (IE),Information laws and regulations (ILR). Specific behaviors include: Healthy and correct use of learning resources to create a good information learning environment (IE1); Restrain one's own information ethical behavior and supervise others' information behavior (IE2); Abide by the network civilization convention, purify the network language, civilized and polite learning and communication (IE3); Impart knowledge of laws, regulations and ethics related to technology utilization (ILR1); Be clear about equal access and access to information and respect for others' intellectual property rights (ILR2).

C. RESEARCH METHOD

According to the general process of learning analysis and machine learning, this study mainly includes data preprocessing, feature extraction, algorithm selection, model training, performance evaluation and result analysis.

The main technical route of this study is shown in Figure 1:

(1)The correlation between learning behavior characteristics and learning effect is calculated on the basis of data cleaning; Observe and analyze the relationship between predictive variables and learning effect, and establish the feature subset participating in model construction.

(2) The five models are trained and tested using the cross-validation method of ten-fold.

(3) On the basis of a single prediction model, the prediction effect of the model is improved by optimizing the algorithm parameters several times.

(4) Carry out prediction effect evaluation and comparative analysis, and establish the optimal prediction algorithm model.

D. DATA COLLECTION AND PREPROCESSING

1) DATA COLLECTION

The research data comes from the "Special Survey on Information Literacy of College Students" implemented by the research group of the "Research on Information Literacy of College Students Supported by Smart Campus", a teaching quality project in Anhui Province, China, in 2022. The study takes into account the impact of scattered, random and representative data on the student population, involving students from a variety of disciplinary and professional backgrounds.The data were collected from the information literacy learning behavior questionnaire data and information literacy course performance data of 320 junior students in Huainan Normal University in 2020. Data was collected by means of a web-based questionnaire administered in batches to students in each class. A pre-survey was conducted before the questionnaire was distributed to test whether the questions were fully understood by the subjects, whether the expression was appropriate and the degree of cooperation, so the overall recall quality of the questionnaire was very high. The data presents positive distribution, with little difference between the data, and good reliability and validity.

2) DATA PREPROCESSING

Figure 2 shows a descriptive statistical overview.The horizontal axis represents variables, and the vertical axis represents numerical values. It gives some indication of the data results for each variable.The descriptive statistics revealed a few missing values and outliers.The Min, Max, Average and Deviation of each feature subset are shown in the figure. Average ranks in the top 3 for IAC1, IPC3 and IPC2, while IB4, IB2 and IB3 rank in the bottom 3; IE1, IPC1 and IAC1 rank in the bottom 3 for Deviation, while IB4, ISK2 and ISK3 rank in the top 3.

To ensure the quality of the classification learning model construction, data preprocessing was performed on the collected learning behavior characteristics and performance data, including operations such as missing value processing, abnormal data processing and data transformation.Outlier processing. SPSS was used to remove the null data and other abnormal data present in the training set. Then the outliers of the data are removed by box plot.After data cleaning, 315 recorded data were finally retained.

Data transformation. In order to enable the machine learning model to achieve better recognition, the data in the collected training set needs to be transformed operationally. The attribute of learning effect should be defined as "nominal attribute". The transformation level attributes are: "5=Excellent", "4=Good", "3=Medium", "2=Pass", "1=Fail".

IV. RESULTS AND DISCUSSION

A. CORRELATION ANALYSIS OF LEARNING BEHAVIOR CHARACTERISTICS AND LEARNING EFFECT

Modeling feature subset selection can be achieved through correlation analysis of learning behavior characteristics and learning effect. Correlation analysis is the analysis of two or more elements of variables that are related as a measure of their degree of association. The related elements must have some kind of association or likelihood in order for correlation analysis to be performed.

If two variables have a strong interdependence, then we can say that the two variables have a high correlation. If the values of both groups increase at the same time, they are said to be positively correlated; if the value of one group increases, then the value of the other group decreases, which is called

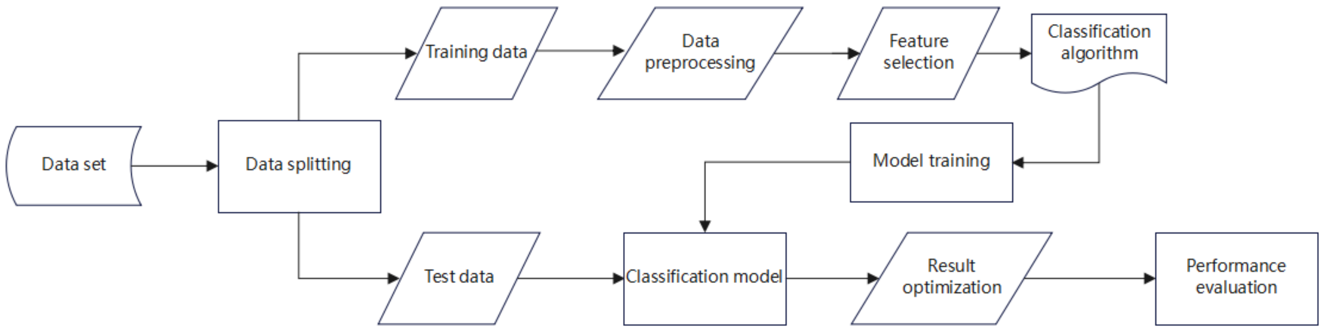

FIGURE 1. Technical road map.

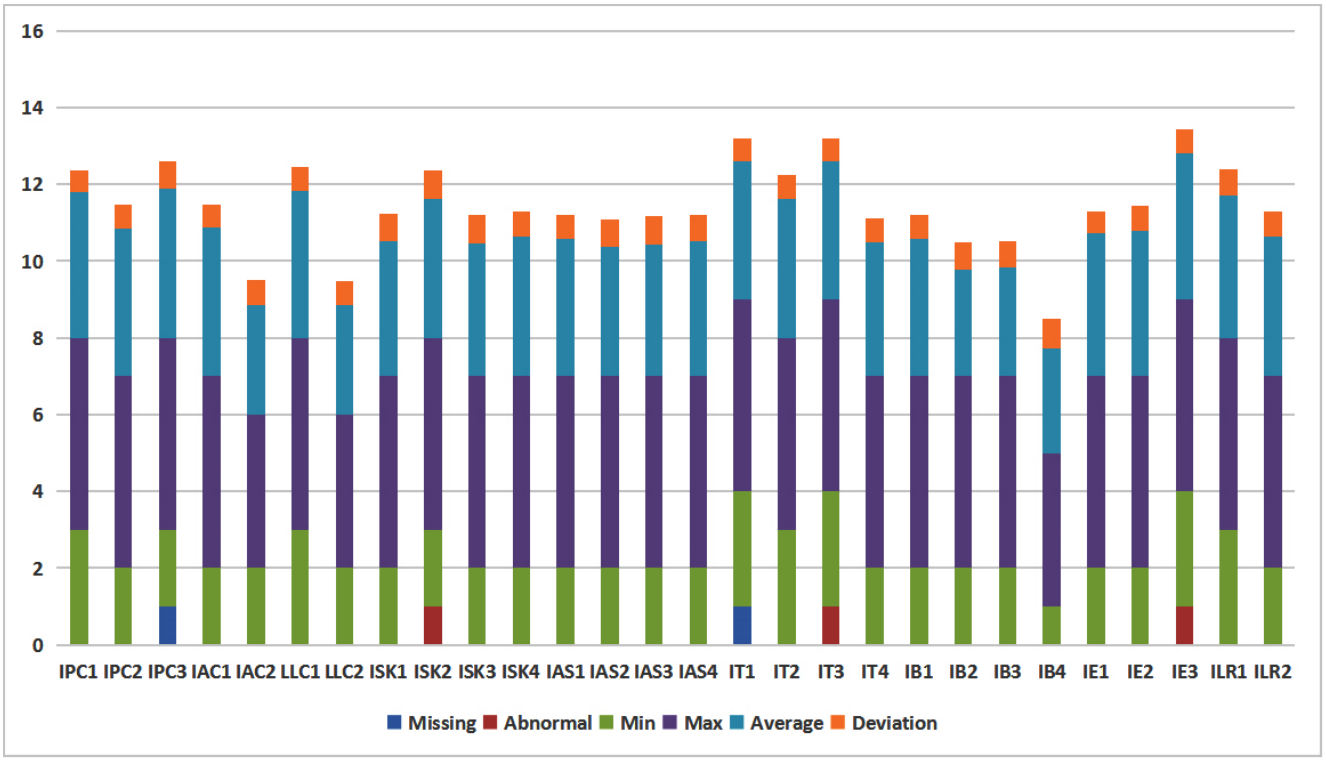

FIGURE 2. Dataset descriptive statistics.

a negative correlation. Pearson's algorithm is used here to calculate the correlation. Pearson's correlation coefficient is an important measure of the interrelationship between two variables, and it has a correlation between -1 and 1. If there are P related variables and the correlation coefficient of the two variables needs to be found, the number of correlation coefficients obtained is as follows:

$$R_{P\times P} = p(p-1)/2 \quad (1)$$

If the variables are arranged into a numerical square in order of their numbering, this square is the correlation matrix.There are two identical variables on the diagonal from the top left to the bottom right, both of which have a value of 1; the correlation coefficient above the diagonal has a symmetric relationship with the part below.

The Pearson correlation coefficient between each variable and the learning effect was calculated to measure the linear correlation between the existing variables. The correlation coefficients between the variables are shown in Figure 4. The intersection of the two variables in the rows and columns is the significance plot, and the color knob at the bottom corresponds to the correlation coefficient. The correlation between the predictor variables and the learning effect is shown in Figure 3. R takes values between -1 and +1. If r>0, it means that the two variables are positively correlated, i.e., the larger the value of one variable, the larger the value of the other variable; if r<0, it means that the two variables are negatively

Attributes	Gender	IPC1	IPC2	IPC3	IAC1	IAC2	LLC1	LLC2	ISK1	ISK2	ISK3	ISK4	IAS1	IAS2	IAS3	IAS4	IT1	IT2	IT3	IT4	IB1	IB2	IB3	IB4	IE1	IE2	IE3	ILR1	ILR2	score
Gender	1	-0.0...	-0.104	-0.1...	-0.123	-0.161	-0.115	0.012	-0.032	-0.147	-0.171	-0.185	-0.120	-0.063	-0.049	-0.108	-0.112	-0.088	0.096	-0.141	-0.166	-0.061	0.103	-0.048	-0.023	-0.061	-0.044	-0.174	0.040	-0.103
IPC1	-0.022	1	0.585	0.476	0.451	0.244	0.368	0.211	0.253	0.195	0.227	0.322	0.320	0.258	0.339	0.257	0.267	0.211	0.354	0.217	0.257	0.272	0.322	0.101	0.164	0.210	0.273	0.305	0.213	0.486
IPC2	-0.104	0.585	1	0.587	0.645	0.500	0.449	0.282	0.288	0.366	0.290	0.364	0.397	0.274	0.340	0.394	0.437	0.335	0.430	0.442	0.337	0.410	0.262	0.211	0.432	0.364	0.456	0.400	0.265	0.647
IPC3	-0.169	0.476	0.587	1	0.659	0.546	0.592	0.327	0.211	0.275	0.171	0.336	0.276	0.148	0.231	0.368	0.346	0.333	0.338	0.360	0.247	0.316	0.186	0.121	0.370	0.396	0.498	0.500	0.343	0.584
IAC1	-0.123	0.451	0.645	0.659	1	0.754	0.611	0.390	0.294	0.390	0.304	0.383	0.425	0.257	0.319	0.428	0.443	0.377	0.359	0.369	0.410	0.299	0.256	0.175	0.470	0.369	0.461	0.479	0.254	0.673
IAC2	-0.161	0.244	0.500	0.546	0.754	1	0.533	0.287	0.271	0.362	0.205	0.296	0.265	0.213	0.273	0.298	0.345	0.199	0.221	0.317	0.253	0.211	0.079	0.134	0.395	0.350	0.328	0.309	0.113	0.525
LLC1	-0.115	0.368	0.449	0.592	0.611	0.533	1	0.720	0.324	0.331	0.285	0.369	0.333	0.229	0.314	0.427	0.394	0.340	0.297	0.351	0.297	0.146	0.205	0.138	0.419	0.356	0.447	0.484	0.335	0.617
LLC2	0.012	0.211	0.282	0.327	0.390	0.287	0.720	1	0.174	0.188	0.253	0.280	0.309	0.255	0.174	0.335	0.257	0.312	0.314	0.312	0.252	0.127	0.243	0.158	0.312	0.242	0.390	0.318	0.265	0.484
ISK1	-0.032	0.253	0.288	0.211	0.294	0.271	0.324	0.174	1	0.485	0.464	0.376	0.455	0.439	0.660	0.422	0.407	0.282	0.285	0.372	0.276	0.255	0.289	0.265	0.255	0.176	0.151	0.270	0.064	0.557
ISK2	-0.147	0.195	0.366	0.275	0.390	0.362	0.331	0.188	0.485	1	0.371	0.319	0.385	0.361	0.345	0.350	0.354	0.251	0.258	0.340	0.344	0.161	0.279	0.224	0.364	0.247	0.298	0.312	0.238	0.536
ISK3	-0.171	0.227	0.290	0.171	0.304	0.205	0.285	0.253	0.464	0.371	1	0.486	0.520	0.758	0.482	0.471	0.498	0.495	0.384	0.461	0.542	0.359	0.287	0.416	0.382	0.345	0.360	0.373	0.278	0.655
ISK4	-0.185	0.322	0.364	0.336	0.383	0.296	0.369	0.280	0.376	0.319	0.486	1	0.689	0.358	0.403	0.744	0.592	0.507	0.482	0.560	0.471	0.411	0.326	0.373	0.414	0.352	0.393	0.455	0.213	0.695
IAS1	-0.120	0.320	0.397	0.276	0.425	0.265	0.333	0.309	0.455	0.385	0.520	0.689	1	0.465	0.561	0.681	0.728	0.509	0.594	0.605	0.547	0.453	0.411	0.387	0.502	0.435	0.445	0.552	0.249	0.766
IAS2	-0.063	0.258	0.274	0.148	0.257	0.213	0.229	0.255	0.439	0.361	0.758	0.358	0.465	1	0.425	0.342	0.426	0.412	0.391	0.441	0.485	0.332	0.285	0.338	0.337	0.248	0.263	0.263	0.129	0.589
IAS3	-0.049	0.339	0.340	0.231	0.319	0.273	0.314	0.174	0.660	0.345	0.482	0.403	0.561	0.425	1	0.372	0.541	0.368	0.409	0.535	0.400	0.368	0.281	0.333	0.351	0.268	0.240	0.330	0.101	0.626
IAS4	-0.108	0.257	0.394	0.368	0.428	0.298	0.427	0.335	0.422	0.350	0.471	0.744	0.681	0.342	0.372	1	0.576	0.466	0.455	0.513	0.413	0.397	0.278	0.318	0.411	0.437	0.495	0.515	0.254	0.704
IT1	-0.112	0.267	0.437	0.346	0.443	0.345	0.394	0.257	0.407	0.354	0.498	0.592	0.728	0.426	0.541	0.576	1	0.609	0.605	0.757	0.626	0.533	0.468	0.517	0.529	0.562	0.535	0.663	0.360	0.810
IT2	-0.088	0.211	0.335	0.333	0.377	0.199	0.340	0.312	0.282	0.251	0.495	0.507	0.509	0.412	0.368	0.466	0.609	1	0.658	0.615	0.568	0.776	0.574	0.512	0.451	0.344	0.431	0.512	0.299	0.722
IT3	0.096	0.354	0.430	0.338	0.359	0.221	0.297	0.314	0.285	0.258	0.384	0.482	0.594	0.391	0.409	0.455	0.605	0.658	1	0.558	0.516	0.572	0.729	0.367	0.499	0.386	0.482	0.449	0.264	0.719
IT4	-0.141	0.217	0.442	0.360	0.369	0.317	0.351	0.312	0.372	0.340	0.461	0.560	0.605	0.441	0.535	0.513	0.757	0.615	0.558	1	0.633	0.477	0.370	0.650	0.524	0.464	0.463	0.565	0.228	0.768
IB1	-0.166	0.257	0.337	0.247	0.410	0.253	0.297	0.252	0.276	0.344	0.542	0.471	0.547	0.485	0.400	0.413	0.626	0.568	0.516	0.633	1	0.470	0.432	0.420	0.572	0.396	0.514	0.467	0.460	0.699
IB2	-0.061	0.272	0.410	0.316	0.299	0.211	0.146	0.127	0.255	0.161	0.359	0.411	0.453	0.332	0.368	0.397	0.533	0.776	0.572	0.477	0.470	1	0.470	0.412	0.373	0.286	0.349	0.378	0.178	0.617
IB3	0.103	0.322	0.262	0.186	0.256	0.079	0.205	0.243	0.289	0.279	0.287	0.326	0.411	0.285	0.281	0.278	0.468	0.574	0.729	0.370	0.432	0.470	1	0.318	0.305	0.223	0.297	0.342	0.258	0.557
IB4	-0.048	0.101	0.211	0.121	0.175	0.134	0.138	0.158	0.265	0.224	0.416	0.373	0.387	0.338	0.333	0.318	0.517	0.512	0.367	0.650	0.420	0.412	0.318	1	0.284	0.283	0.220	0.330	0.101	0.531
IE1	-0.023	0.164	0.432	0.370	0.470	0.395	0.419	0.312	0.255	0.364	0.382	0.414	0.502	0.337	0.351	0.411	0.529	0.451	0.499	0.524	0.572	0.373	0.305	0.284	1	0.543	0.669	0.540	0.375	0.682
IE2	-0.061	0.210	0.364	0.396	0.369	0.350	0.356	0.242	0.176	0.247	0.345	0.352	0.435	0.248	0.268	0.437	0.562	0.344	0.386	0.464	0.396	0.286	0.223	0.283	0.543	1	0.734	0.604	0.480	0.611
IE3	-0.044	0.273	0.456	0.498	0.461	0.328	0.447	0.390	0.151	0.298	0.360	0.393	0.445	0.263	0.240	0.495	0.535	0.431	0.482	0.463	0.514	0.349	0.297	0.220	0.669	0.734	1	0.683	0.664	0.681
ILR1	-0.174	0.305	0.400	0.500	0.479	0.309	0.484	0.318	0.270	0.312	0.373	0.455	0.552	0.263	0.330	0.515	0.663	0.512	0.449	0.565	0.467	0.378	0.342	0.330	0.540	0.604	0.683	1	0.484	0.711
ILR2	0.040	0.213	0.265	0.343	0.254	0.113	0.335	0.265	0.064	0.238	0.278	0.213	0.249	0.129	0.101	0.254	0.360	0.299	0.264	0.228	0.460	0.178	0.258	0.101	0.375	0.480	0.664	0.484	1	0.430
score	-0.103	0.486	0.647	0.584	0.673	0.525	0.617	0.484	0.557	0.536	0.655	0.695	0.766	0.589	0.626	0.704	0.810	0.722	0.719	0.768	0.699	0.617	0.557	0.531	0.682	0.611	0.681	0.711	0.430	1

FIGURE 3. Correlation matrix data.

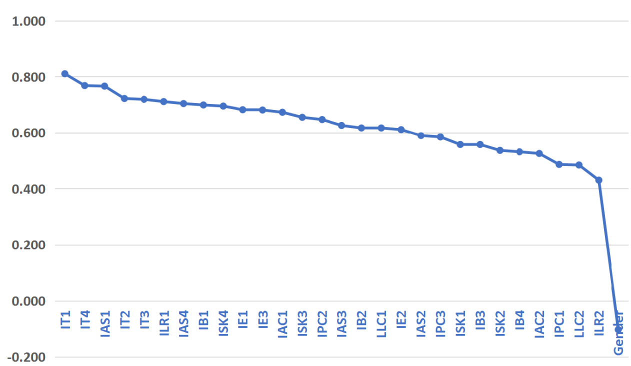

FIGURE 4. Correlation between predictor variables and learning effect.

correlated, i.e., the larger the value of one variable, the smaller the value of the other variable. The larger the absolute value of r, the stronger the correlation; the smaller the absolute value of r, the weaker the correlation [42].The linear correlation between the existing variables was measured by calculating Pearson's correlation coefficients between the variables of information literacy of college students and learning effect. It was concluded that the vast majority of the predictor variables showed some positive correlation with learning outcomes. As the correlations reflect some variability, this provides support for the analysis of learning behavioural characteristics.

1) ANALYSIS OF HIGH CORRELATION LEARNING BEHAVIOR CHARACTERISTICS

The top five variables are IT1, IT4, IAS1, IT2, and IT3, and it can be seen that the relationship between Information thinking (IT) and learning effect in Application and innovation is the most significant. The top three variables with the highest correlation with learning effect were IT1 with 0.810, IT4 with 0.768, and IAS1 with 0.766.

IT1 is defining and identifying implicit assumptions in information and inferring information. IT1 is concerned with information mining and targeted critical cognition. IT4 is constructing problem solutions by integrating resources and using rational algorithms. IT4 is concerned with information creation, focusing on specific solutions to problems and emphasizing the development of creativity. ias1 is using various search engines and web platforms to find the required information. It is an information application skill. This learning behavior focuses on the ability of college students to access information. The results of the analysis suggest the development of application and innovation skills, and the increasing of information acquisition skills in knowledge and skills.Integrating critical thinking methods into the information literacy education system Zhuozhuo et al. [43] is an important strategy to improve the innovation and creativity ability of college students and break through the bottleneck of information literacy education. At the same time, attention should still be paid to cultivating and improving students' information literacy knowledge and skills, especially the improvement of information acquisition ability in the information age.

2) ANALYSIS OF LOW CORRELATION LEARNING BEHAVIOR CHARACTERISTICS

The three variables with the lowest correlations with learning effect were ILR2 with 0.430, LLC2 with 0.484, and IPC1 with 0.486. ILR2 is learning equal access to information and respecting the intellectual property rights of others; LLC2 is using information technology to support professional and personal development; IPC1 is identifying and classifying information. From the perspective of learning behavior presentation space, these three learning behaviors are abstract in nature; from the perspective of learning behavior presentation time, these three learning behaviors are less integrated with college students' study, life and existing learning environment than other learning behavior characteristics. This suggests a reference for later pedagogical improvement, which requires more streamlined and effective learning behaviors for college students in response to these indicators.

3) ESTABLISHMENT OF A SUBSET OF LEARNING BEHAVIOR CHARACTERISTICS

In order to construct better prediction models and achieve better prediction results with fewer features, three learned behavioral features with correlations below 0.500 were not involved in the prediction model construction, namely IPC1 (0.486), LLC2 (0.484), and ILR2 (0.430). The gender variable was also not involved in the prediction model construction because its degree of relationship with learning effect was -0.103.

To sum up, a subset of 25 information literacy learning behavior characteristics of college students is currently retained, specifically IT1, IT4, IAS1, IT2, IT3, ILR1, IAS4, IB1, ISK4, IE1, IE3, IAC1, ISK3, IPC2, IAS3, IB2, LLC1, IE2, IAS2, IPC3, ISK1, IB3, ISK2, IB4, IAC2.

B. CLASSIFICATION MODEL PREDICTION OF LEARNING EFFECT

The performance evaluation metrics of the binary classification model include Accuracy, Precision, Recall, F1-Score (F1), etc. [44].TP indicates the number of positive samples whose learning effect was correctly predicted; TN indicates the number of negative samples whose learning effect was correctly predicted; FP indicates the number of samples that were incorrectly predicted as positive; and FN indicates the number of samples that were incorrectly predicted as negative.

Classification Accuracy is the percentage of the number of correct samples that can be predicted in the classification model and reflects the accuracy of the overall classification.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (2)$$

The Precision is the ratio of the number of positive cases correctly predicted by the classification model to the number of all positive cases predicted by the classification model, i.e., the proportion of true positive cases among all results predicted as positive.

$$Precision = TP/(TP + FP) \quad (3)$$

The Recall is the ratio of the number of positive samples correctly predicted by the classification model to the actual number of positive samples in the entire test set, i.e., the proportion of true positive cases that are found by the classification model.

$$Recall = TP/(TP + FN) \quad (4)$$

The F1-Score is a comprehensive metric that combines Precision and Recall. Since Precision and Recall are a pair of contradictory measures, and different problems focus on different criteria, F1-Score is a good comprehensive evaluation metric, and the larger the value of this metric, the better.

$$F1 - Score = 2 \times Precision \times Recall/(Precision + Recall) \quad (5)$$

The kappa (KIA) coefficient is a measure of classification accuracy. kia is an index that enables the calculation of overall consistency and classification consistency. The KIA is used to perform an assessment of the accuracy of a multi-classification model. The higher the value of this coefficient, the higher the classification accuracy achieved by the model. kappa coefficient can be calculated as follows. Po denotes the proportion of observation accuracy or consistency cells. Pc indicates the proportion of cells that are contingently consistent or expected to be contingently consistent.

$$KIA = (Po - Pc)/(1 - Pc) \quad (6)$$

In terms of prediction model selection, the main focus is on predicting college students' learning effect levels through their information literacy behavioral performance indicators. This is a typical classification problem, so the classical machine learning classification algorithm is used to compare the prediction performance of different models separately for this study sample. In the following, the five models are trained and tested using a ten-fold cross validation method. The dataset is first divided into ten parts, then nine of them are rotated as training data and the remaining one is used as test data, and finally the model training is performed by maximizing the use of samples by averaging the correct rate each time as the evaluation value of the algorithm accuracy.

1) DECISION TREE

Decision tree is a greedy algorithm that classifies instances based on features and performs recursive binary partitioning on the feature space. Starting from the root node of the tree, the sample data is compared with the feature nodes in the decision tree, and the branches at the next level are selected to continue the comparison based on the judgment result,

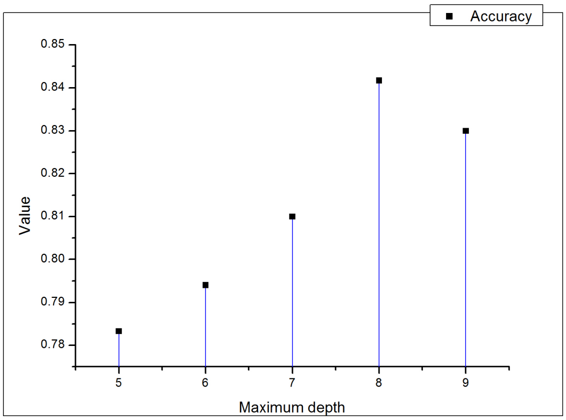

FIGURE 5. The selection process of Decision Tree hyperparameters.

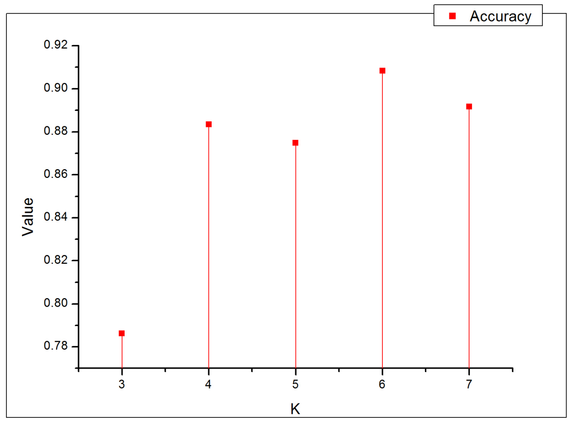

FIGURE 6. The selection process of KNN hyperparameters.

and the final leaf node is the classification result [45].The advantage of decision trees is that they are more readable and faster to classify [46].The C4.5 decision tree algorithm uses the ''gain ratio'' to select the optimal partitioning attribute.

By training and optimizing the Decision Tree parameters, the best recognition effect of the model is obtained when the core parameter Maximum depth is set to 8, the minimum leaf size is 2 and the confidence is 0.1. The experimental process of obtaining Decision Tree is shown in Figure 5. Observe the results of multiple experiments and get the optimal accuracy rate of 84.17%.

2) K-NEAREST NEIGHBOR

The K-Nearest Neighbor (KNN) algorithm is an algorithm based on statistical classification. The advantage of this algorithm is that it does not need to partition the vector space consisting of all data records, and the classification is better by training the model data to find K similar vectors, and the disadvantage is that it is insensitive to outliers [47].

By training and optimizing the KNN parameters, the model recognition effect is best when the core parameter K is set to 6. The experimental procedure of KNN is obtained as shown in Figure 6. The optimal accuracy rate of 90.83% is obtained by observing the results of multiple experiments.

3) NAIVE BAYES

Naive Bayes is a data detection and classification algorithm based on probability theory. The algorithm can relate the prior and posterior probabilities of events and use sample data with prior information to determine the posterior probability of events. Its advantage is that the model is simple to construct and has high efficiency and stability [48].

By training and optimizing the Naive Bayes parameters, the model recognition effect is best when the core parameter minimum bandwidth is set to 0.2. The experimental procedure of obtaining Naive Bayes is shown in Figure 7. The optimal accuracy rate of 90.00% is obtained by observing the results of multiple experiments.

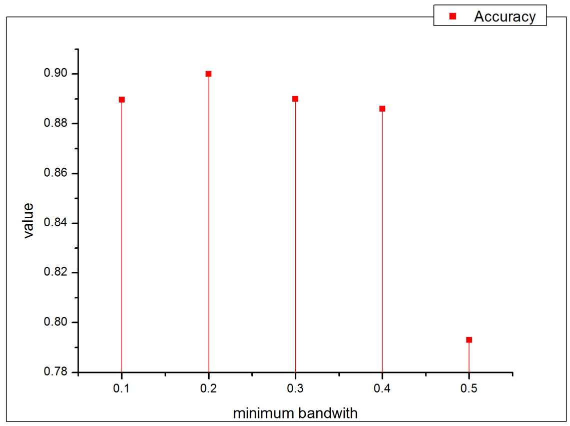

FIGURE 7. The selection process of Naive Bayes hyperparameters.

4) NEURAL NET

Neural network is a mathematical model that simulates biological neural networks for information processing, and neural networks are applied in classification problems with good results [49]. Neural networks are mainly composed of: input layer, hidden layer, and output layer.

By training and optimizing the Neural Net parameters, the best recognition effect of the model is obtained when the core parameters momentum is set to 0.9, training cycles to 200, and learning rate to 0.01. The experimental process of obtaining Neural Net is shown in Figure 8. Observing the results of multiple experiments, the optimal accuracy rate is obtained as 91.67%.

5) RANDOM FOREST

Random Forest utilizes random sampling of data samples and features to train multiple tree classifiers, avoiding the learning of all samples and all features per tree, thus increasing randomness, avoiding overfitting, and integrating the results of a single decision tree according to the rules of Bagging [50].The training sample data are sampled with put-back to generate K classification regression trees; assume that there are n features in the feature space and m features are

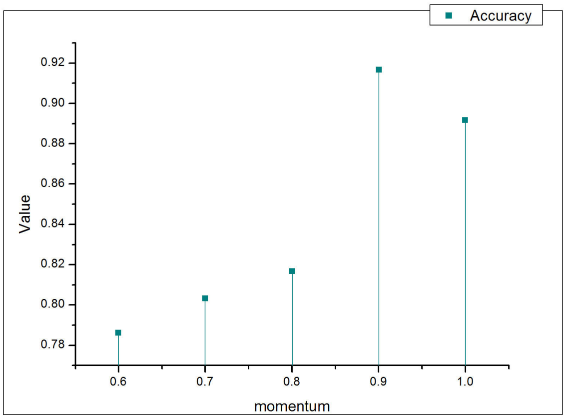

FIGURE 8. The selection process of Neural Net hyperparameters.

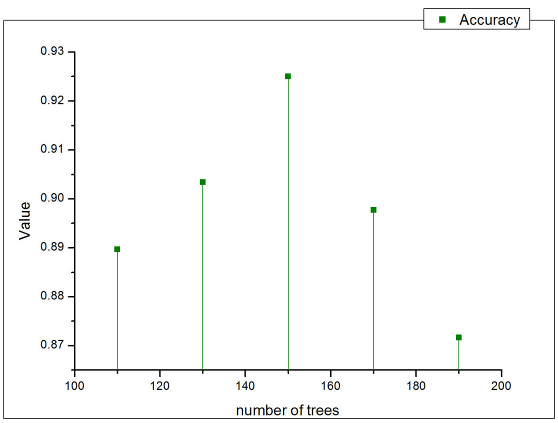

FIGURE 9. The selection process of Random Forest hyperparameters.

TABLE 5. Prediction results of the classification model.

Model	Accuracy	Precision	Recall	F1-Score	Kapaa
Decision Tree	84.17%	76.00%	88.63%	81.83%	0.703
KNN	90.83%	80.87%	93.96%	86.92%	0.831
Naive Bayes	90.00%	93.06%	69.97%	79.88%	0.806
Neural Net	91.67%	60.83%	63.15%	63.11%	0.837
Random Forest	92.50%	84.56%	94.81%	89.39%	0.859

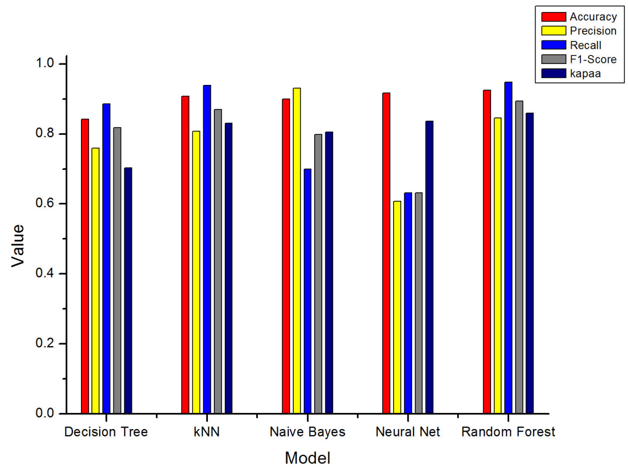

FIGURE 10. Performance comparison of learning effect prediction models.

randomly selected at the nodes of each tree, requiring m < n; make each tree grow maximally without any pruning; form a forest by multiple trees, and the classification results are determined by how many tree classifiers vote.

By training and optimizing the parameters of the Random Forest model, the best recognition effect of the model was obtained when the number of trees parameter was set to 150 and the criterion was set to gain_ratio. The experimental results of the Random Forest are shown in Figure 9 after repeated execution for several times. Observe the results of the multiple experiments and get the optimal accuracy rate of 92.50%.

After the parameter tuning of each model, the prediction results of each model are obtained as shown in Table 5. The visual illustration of the prediction models is shown in Figure 10. The range of kappa taking values represents different degrees. 0.1∼0.2: slight; 0.2∼0.4: fair; 0.4∼0.6: moderate; 0.6∼0.8: substantial; 0.8∼1.0: almost perfect [51]. The kappa of each model kappa indicates that the overall consistency and classification consistency of each model are normal and basically meet the requirements.

From the analysis of the prediction results, the highest Accuracy is Random Forest, reaching 92.50%, followed by Neural Net, KNN; the highest Precision is Naive Bayes, reaching 93.06%, followed by Random Forest, KNN; the highest Recall is Random, reaching 94.81%, followed by KNN, Decision Tree; the highest F1-Score is Random Forest, reaching 89.39%, followed by KNN, Naive Bayes; the highest Kapaa is Random Forest, reaching 0.859, followed by Neural Net The results of all indicators show that Random Forest prediction model has the best performance and can be used to enhance the learning effect prediction of college students' information literacy.

C. DISCUSSION

Machine learning research is not simply to pursue the high accuracy of the machine learning prediction model, but more importantly to explore the characteristics that can be explained and trusted for optimizing the teaching process [52].Through the correlation study, the intervention measures needed to improve the teaching quality of college students' information literacy are clarified. Pearson algorithm is used to analyze the learning behavior characteristics of college students' information literacy, revealing that there is a more significant correlation between information thinking and information application skills and learning effect, especially information thinking.Information literacy education should focus on cultivating students' innovative spirit and practical ability.Some universities try to cultivate critical thinking from the perspective of knowledge transfer [53]; some scholars propose the integration model of information literacy education with professional courses that integrates critical thinking [54], adopt the information literacy education model that integrates innovation and entrepreneurship training of college students to cultivate students' innovative thinking by using the "holistic thinking approach" [55], and improve critical reflection ability based on the "163" information literacy education system of "Internet+" [56].However, there is a lack of a complete training system for critical thinking ability and innovative thinking ability.

Combining the four metrics of accuracy, precision, recall, F1 value, and kappa, the performance of the classifier obtained by the Random Forest algorithm is optimal in terms of prediction performance for all types of models. The higher prediction accuracy proves the effect of machine learning algorithms applied to learning effect prediction modeling, which is basically in line with the findings of related studies such as Juan et al. [57] and Faqin et al. [58]. By using the Random Forest algorithm model for predicting the learning effect of college students' information literacy, we can predict the learning effect of college students in information literacy education more accurately, guide the adjustment of teaching behaviors and allocation of teaching resources, and effectively guarantee teaching quality.

V. CONCLUSION

The results prove that the prediction model proposed in this paper has a significant effect on the cultivation of information literacy of college students. On the one hand, Algorithmic analysis of the learning behaviour characteristics of college students' information literacy reveals a more significant correlation between information thinking, information application skills and learning effect. Emphasis should be placed on the cultivation of information thinking, while not neglecting the cultivation of information acquisition ability. Universities should understand the importance and urgency of information literacy education for college students from the height of sustainable development of talents and cultivation of high-quality and innovative talents. It is necessary to make full use of network and multimedia technologies to provide intelligent learning tools and learning environments conducive to independent, cooperative and research learning, to incorporate critical thinking methods into the information literacy education system, to establish a long-term assessment mechanism oriented to the cultivation of critical thinking, and to actively promote the critical thinking cognition and knowledge creation skills of college students. On the other hand, it is necessary to further optimize the learning methods in such aspects as information laws and regulations, streamline the learning contents in such aspects as information perception awareness, and continuously instill and guide college students to establish the concept of lifelong learning. This study provides a more reliable data base for educational administrators to analyze the potential connections between information literacy education phenomena and outcomes, thereby increasing the success rate of educational decisions.

In conclusion, this study proposes an effective machine learning approach to characterize the learning behaviors and predict the learning effects of information literacy among college students. This study uses a data-driven thinking to promote teachers and students to optimize their learning paths and improve the effectiveness of information literacy instruction. It also strongly supports the implementation of differentiated teaching decision-making [59] and the construction of a long-term mechanism for differentiated educational decisions through a data-driven approach.

Although this study has conducted some exploration, there are still limitations. While machine learning approaches work to some extent, the study suggests that technological tools need to be applied according to the specific teaching and learning situation. Further research is proposed in the following areas at a later stage.(1)Learning effect prediction has not been able to cover other possible factors in learning scenarios, which puts higher demands on the quality of the learning behavior trait scale. We will explore the learning behavior characteristics in more scenarios, evolve the learning behavior characteristics scale, improve the universality, and form a closed loop of textbook development of teaching experiment, teaching research, and teaching practice.(2)In this study, only five supervised classification algorithms are used, such as Decision Tree, KNN, Naive Bayes, Neural Net, and Random Forest. In subsequent studies, the adopted algorithms can be collectively improved to achieve better prediction results.

ACKNOWLEDGMENT

The author would like to thank everyone who made every effort to improve the content of this research paper. They also like to thank the anonymous reviewers and editors for their work on the publication of this paper.

REFERENCES

[1] Z. Changhai, "Research on the information literacy education model of Chinese college students based on critical thinking and creativity," *J. China Library*, vol. 4, no. 15, pp. 15–16, Aug. 2016, doi: 10.13530/j.cnki.jlis.164008.

[2] S. Hui, "Information literacy education for college students," *Educ. Theory Pract.*, vol. 30, no. 10, pp. 38–39, Oct. 2008.

[3] D. Yan and G. Li, "A heterogeneity study on the effect of digital education technology on the sustainability of cognitive ability for middle school students," *Sustainability*, vol. 15, no. 3, pp. 2784–2786, Feb. 2023, doi: 10.3390/su15032784.

[4] Y. Xiaohong, Z. Xin, and Z. Jing, "Research on the Chinese experience of the 'epidemic' of the education war—Online education perspective," *China Distance Educ.*, vol. 43, no. 1, pp. 1–11, Jan. 2023, doi: 10.13541/j.cnki.chinade.2023.01.002.

[5] G. Shan and F. Shan, "Digital literacy practice and enlightenment of the University of Queensland library in Australia," *Library Sci. Res.*, vol. 7, no. 10, pp. 95–101, Jul. 2019, doi: 10.15941/j.cnki.issn1001-0424.2019.10.015.

[6] T. H. Wufati, "Mining the characteristics of meaningful learning behavior: A framework for predicting learning outcomes," *Open Educ. Res.*, vol. 25, no. 6, pp. 75–82, Jan. 2019, doi: 10.13966/j.cnki.kfjyyj.2019.06.008.

[7] H. Hang, L. Yaxin, L. Qi'e, Y. Hairu, Z. Qiuhua, and C. Yifan, "The occurrence process, design model and mechanism explanation of deep learning," *China's Distance Educ.*, vol. 5, no. 1, pp. 54–61, Mar. 2020, doi: 10.13541/j.cnki.chinade.2020.01.005.

[8] I. A. AlShammari, M. Aldhafiri, and Z. Al-Shammari, "A meta-analysis of educational data mining on improvements in learning outcomes," *College student J.*, vol. 47, no. 2, pp. 326–333, Jun. 2013.

[9] W. Gaihua and F. Gangshan, "Prediction of online learning behavior and achievement and design of learning intervention model," *China's Distance Educ.*, vol. 2, no. 2, pp. 39–48, Mar. 2019, doi: 10.13541/j.cnki.chinade.20181214.007.

[10] *UNESCO*. Accessed: Apr. 1, 2023. [Online]. Available: https://www.unesco.org/en/articles/challenges-and-opportunities-artificial-intelligence-education

[11] G. Yang, B. Wen, and W. Lin, "Research status, hot spots and enlightenment of college students' information literacy: Based on bibliometric analysis of CNKI from 2000 to 2021," in *Proc. 4th World Symp. Softw. Eng.*, Sep. 2022, pp. 161–166, doi: 10.1145/3568364.3568389.

[12] J. Li, "Machine learning-based evaluation of information literacy enhancement among college teachers," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 22, pp. 116–131, Nov. 2022, doi: 10.3991/ijet.v17i22.35117.

[13] L. Yu, D. Wu, H. H. Yang, and S. Zhu, "Smart classroom preferences and information literacy among college students," *Australas. J. Educ. Technol.*, vol. 38, no. 2, pp. 144–163, Feb. 2022, doi: 10.14742/ajet.7081.

[14] Y. Ying, "Research on college students' information literacy based on big data," *Cluster Comput.*, vol. 22, no. S2, pp. 3463–3470, Mar. 2019, doi: 10.1007/s10586-018-2193-0.

[15] Z. Mian, Y. Bai, and R. K. Ur, "Research on college computer-computing and information literacy online course based on MOOC: Taking the north minzu university as an example," in *Proc. IEEE 3rd Int. Conf. Comput. Sci. Educ. Inf. (CSEI)*, Jun. 2021, vol. 6, no. 7, pp. 300–306, doi: 10.1109/CSEI51395.2021.9477751.

[16] M. S. Haider and C. Ya, "Assessment of information literacy skills and information-seeking behavior of medical students in the age of technology: A study of Pakistan," *Inf. Discovery Del.*, vol. 49, no. 1, pp. 84–94, Feb. 2021, doi: 10.1108/IDD-07-2020-0083.

[17] X. Ouyang, Y. Xiao, and J. Zhong, "Research on the influencing factors and the promotion measures of college students' information literacy," in *Proc. 8th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Dec. 2016, vol. 12, no. 12, pp. 728–732, doi: 10.1109/ITME.2016.0169.

[18] T. Nishikawa and G. Izuta, "The information technology literacy level of newly enrolled female college students in Japan," *Hum. Social Sci. Rev.*, vol. 7, no. 1, pp. 1–10, Mar. 2019, doi: 10.18510/hssr.2019.711.

[19] R. Diani, A. Susanti, N. Lestari, M. Saputri, and D. Fujiani, "The influence of connecting, organizing, reflecting, and extending (CORE) learning model toward metacognitive abilities viewed from students' information literacy in physics learning," *J. Phys., Conf. Ser.*, vol. 1796, no. 1, pp. 1–9, Feb. 2021, doi: 10.1088/1742-6596/1796/1/012073.

[20] F. Liu and Q. Zhang, "A new reciprocal teaching approach for information literacy education under the background of big data," *Int. J. Emerg. Technol. Learn*, vol. 16, no. 3, pp. 246–260, Feb. 2021, doi: 10.3991/ijet.v16i03.20459.

[21] M. K. Masko, K. Thormodson, and K. Borysewicz, "Using case-based learning to teach information literacy and critical thinking skills in undergraduate music therapy education: A cohort study," *Music Therapy Perspect.*, vol. 38, no. 2, pp. 143–149, Oct. 2020, doi: 10.1093/mtp/miz025.

[22] R. Xu, C. Wang, Y. Hsu, and X. Wang, "Research on the influence of DNN-based cross-media data analysis on college students' new media literacy," *Comput. Intell. Neurosci.*, vol. 2022, Aug. 2022, Art. no. 9224834, doi: 10.1155/2022/9224834.

[23] M. Aydın, "A multilevel modeling approach to investigating factors impacting computer and information literacy: ICILS Korea and Finland sample," *Educ. Inf. Technol.*, vol. 27, no. 2, pp. 1675–1703, Aug. 2021, doi: 10.1007/s10639-021-10690-1.

[24] Y. Sun, Z. Tan, Z. Li, and S. Long, "Predicting and analyzing college students' performance based on multifaceted data using machine learning," in *Proc. 4th Int. Conf. Adv. Comput. Technol., Inf. Sci. Commun. (CTISC)*, Apr. 2022, pp. 1–6, doi: 10.1109/CTISC54888.2022.9849815.

[25] J. Guo and T. Xu, "IT monitoring and management of learning quality of online courses for college students based on machine learning," *Mobile Inf. Syst.*, vol. 2022, Sep. 2022, Art. no. 5501322, doi: 10.1155/2022/5501322.

[26] Y. Jia and E. Wang, "Research on information anxiety of college students under the background of information overloaded based on support vector machine optimization alogrithm," in *Proc. 2nd Int. Conf. Inf. Sci. Educ. (ICISE-IE)*, Nov. 2021, pp. 484–487, doi: 10.1109/ICISE-IE53922.2021.00117.

[27] C. Pei, "The construction of a prediction model for the teaching effect of two courses education in colleges and universities based on machine learning algorithms," *Wireless Commun. Mobile Comput.*, vol. 2022, Aug. 2022, Art. no. 1167454, doi: 10.1155/2022/1167454.

[28] H. Xu, "GBDT-LR: A willingness data analysis and prediction model based on machine learning," in *Proc. IEEE Int. Conf. Adv. Electr. Eng. Comput. Appl. (AEECA)*, no. 8, Aug. 2022, pp. 396–401, doi: 10.1109/AEECA55500.2022.9919013.

[29] T. Li, "students' numeracy and literacy aptitude analysis and prediction using machine learning," *J. Comput. Commun.*, vol. 10, no. 8, pp. 90–103, Aug. 2022, doi: 10.4236/jcc.2022.108006.

[30] X. Shen and C. Yuan, "A college student behavior analysis and management method based on machine learning technology," *Wireless Commun. Mobile Comput.*, vol. 2021, Aug. 2021, Art. no. 3126347, doi: 10.1155/2021/3126347.

[31] R. Wang, "Research on effectiveness of college English blended teaching mode under small private online course based on machine learning," *Social Netw. Appl. Sci.*, vol. 5, no. 2, pp. 1–13, Feb. 2023, doi: 10.1007/s42452-023-05278-y.

[32] E. Rusdiana, Q. Violinda, C. Pramana, R. Y. Purwoko, D. D. Chamidah, N. Rahmah, Y. J. Prihanto, F. Y. Hasnawati, R. Susanti, A. Y. Haimah, R. Purwahida, F. Arkiang, M. A. Equatora, and A. Mujib, "College students' perception of electronic learning during COVID-19 pandemic in Indonesia: A cross-sectional study," *J. Higher Educ. Theory Pract.*, vol. 10, no. 13, pp. 29–44, Oct. 2022, doi: 10.33423/jhetp.v22i13.5505.

[33] X. Liu and C. Yang, "Analysis of college students' practical teaching effect based on machine learning correlation analysis algorithm: Take the software technology course as an example," in *Proc. 8th Int. Conf. Educ., Manage., Inf. Manage. Soc. (EMIM)*, 2018, vol. 8, no. 1, pp. 782–788, doi: 10.2991/emim-18.2018.159.

[34] H. Akram, A. H. Abdelrady, A. S. Al-Adwan, and M. Ramzan, "Teachers' perceptions of technology integration in teaching-learning practices: A systematic review," *Frontiers Psychol.*, vol. 13, Jun. 2022, Art. no. 920317, doi: 10.3389/fpsyg.2022.920317.

[35] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "student engagement predictions in an e-learning system and their impact on student course assessment scores," *Comput. Intell. Neurosci.*, vol. 2018, Oct. 2018, Art. no. 6347186, doi: 10.1155/2018/6347186.

[36] V. G. Karthikeyan, P. Thangaraj, and S. Karthik, "Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation," *Soft Comput.*, vol. 24, no. 24, pp. 18477–18487, Dec. 2020, doi: 10.1007/s00500-020-05075-4.

[37] H. Hang, D. Shuang, L. Jiarou, and K. Zhonglin, "Construction of learning performance prediction model: From big data analysis of learning behavior," *China's Distance Educ.*, vol. 41, no. 4, pp. 8–20, May 2021, doi: 10.13541/j.cnki.chinade.2021.04.002.

[38] M. Zhijia and W. Fati, "MOOC learning results prediction index exploration and learning group characteristics analysis," *Res. Mod. Distance Educ.*, vol. 147, no. 3, pp. 58–66, Jun. 2017, doi: 10.3969/j.issn.1009-5195.2017.03.008.

[39] *Rapidminer Official Website*. Accessed: Feb. 1, 2023. [Online]. Available: https://rapidminer.com/why-rapidminer/

[40] Q. Dongming, Z. Yumeng, L. Baige, and C. Zhiyun, "Research on the evaluation criteria of college students' information literacy—Taking Shanghai as an example," *China Higher Educ. Res.*, vol. 349, no. 9, pp. 53–59, Oct. 2022, doi: 10.16298/j.cnki.1004-3667.2022.09.10.

[41] Y. Shi, F. Peng, and F. Sun, "A blended learning model based on smart learning environment to improve college students' information literacy," *IEEE Access*, vol. 10, pp. 89485–89498, 2022, doi: 10.1109/ACCESS.2022.3201105.

[42] S. Fang, B. Tang, Z. Fu, C. Luo, and J. Hu, "A study of the factors influencing Pearson's correlation coefficient in the development of higher education—Taking the Yangtze River delta as an example," *Adv. Social Sci.*, vol. 12, no. 01, pp. 49–58, 2023, doi: 10.12677/ASS.2023.121007.

[43] L. Zhuozhuo, M. Miaoer, and Z. Kang, "Cultivating critical thinking in information literacy education—Investigation and experience of foreign university libraries," *Inf. Data Work*, vol. 43, no. 4, pp. 103–112, Sep. 2022, doi: 10.12154/j.qbzlgz.2022.04.011.

[44] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models," *Proc. 1st Workshop Eval. Comparison NLP Syst.*, Nov. 2020, vol. 11, no. 1, pp. 79–91, doi: 10.18653/v1/2020.eval4nlp-1.9.

[45] A. Weipeng and S. Jiaze, "Improvement and analysis of decision tree C4.5 algorithm," *Comput. Eng. Appl.*, vol. 55, no. 12, pp. 169–173, Jun. 2019, doi: 10.3778/j.issn.1002-8331.1805-0482.

[46] Y. Liyuan, "The application of decision tree algorithm in college performance analysis," *J. Weifang Univ.*, vol. 8, no. 4, pp. 57–58, Dec. 2008, doi: 10.3969/j.issn.1671-4288.

[47] Z. Yu, H. Chen, J. Liu, J. You, H. Leung, and G. Han, "Hybrid *k*-nearest neighbor classifier," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1263–1275, Jun. 2015, doi: 10.1109/TCYB.2015.2443857.

[48] M. S. Roobini, K. Babu, J. Joseph, and G. Ieshwarya, "Predicting stock price using data science technique," in *Proc. 2nd Int. Conf. Artif. Intell. Smart Energy (ICAIS)*, no. 2, Feb. 2022, pp. 1013–1020, doi: 10.1109/ICAIS53314.2022.9742772.

[49] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014, doi: 10.5555/2627435.2670313.

[50] C. Zhan, Y. Zheng, H. Zhang, and Q. Wen, "Random-forest-bagging broad learning system with applications for COVID-19 pandemic," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15906–15918, Nov. 2021, doi: 10.1109/JIOT.2021.3066575.

[51] H.-J. Park, S. S. Kim, S.-Y. Lee, N.-H. Park, E.-C. Chung, M.-H. Rho, H.-J. Kwon, and S.-H. Kook, "A practical MRI grading system for cervical foraminal stenosis based on oblique sagittal images," *Brit. J. Radiol.*, vol. 86, no. 1025, May 2013, Art. no. 20120515, doi: 10.1259/bjr.20120515.

[52] L. Sannvya, L. Shengyingjie, and S. Jianwen, "Several key issues in the development of intelligent education," *China's Distance Educ.*, vol. 555, no. 4, pp. 1–7, May 2021, doi: 10.13541/j.cnki.chinade.2021.04.001.

[53] S. Jinfeng, "Critical thinking education and enlightenment at Stanford University," *Higher Educ. Forum*, vol. 259, no. 5, pp. 122–124, Jul. 2021, doi: 10.3969/j.issn.1671-9719.

[54] Z. Changhai, "Research on the information literacy education model of Chinese college students based on critical thinking and creativity," *J. China Library*, vol. 42, no. 4, pp. 102–116, Apr. 2016, doi: 10.13530/j.cnki.jlis.164008.

[55] H. Hongmei, "Research on the optimization of information literacy training mode in colleges and universities under the environment of 'innovation,'" *Library Construct.*, vol. 296, no. 2, pp. 142–147, Feb. 2019, doi: 10.19764/j.cnki.tsgjs.20182378.

[56] J. Haixia, "Research on the construction of information literacy education system based on 'internet plus 'thinking,'" *Mod. Inf.*, vol. 37, no. 2, p. 5, Mar. 2017, doi: 10.3969/j.issn.1008-0821.

[57] W. Juan, Z. Dongwei, L. Falun, Z. Xianglong, and X. Wei, "Analysis of academic behavior and performance prediction of online learners in open education based on data mining," *China Adult Educ.*, vol. 13, no. 3, pp. 29–33, Sep. 2022, doi: 10.3969/j.issn.1004-6577.

[58] S. Faqin and F. Rui, "Research on the influencing factors of online academic achievement based on learning analysis," *China's Audio-Vis. Educ.*, vol. 24, no. 3, pp. 48–54, Apr. 2019, doi: 10.3969/j.issn.1006-9860.

[59] Z. Xuebo and L. Shubing, "Data-driven differentiated teaching decision-making: Crux, logic and mechanism," *Res. Mod. Distance Educ.*, vol. 34, no. 3, pp. 48–57, May 2022, doi: 10.3969/j.issn.1009-5195.

YONG SHI was born in 1984. He received the master's degree from Anhui Normal University. He is currently pursuing the Ph.D. degree in information technology with St. Paul University Philippines. He is a Teacher with the School of Computer Science, Huainan Normal University, a Certified MOOC Teacher with the China School Online, an Innovation and Entrepreneurship Instructor for college students in Anhui, and an advanced individual with the Anhui Education and Research Network. He has five projects with the national and provincial level, guided college students to participate in various competitions and won more than 30 awards at the national provincial level. He has published more than 20 articles, and has a number of soft works and patents. His research interests include information education and data mining.

FANG SUN was born in October 1978. He received the master's degree in educational technology from Beijing Normal University. He is currently an Associate Professor with the School of Education, Huainan Normal University. He is the author of *The Essence of Flash Courseware Making Technology* and other books and founded the influential educational technology Forum (www.etthink.com) and other resources (www.qiexing.com). His research interests include information teaching, application of educational technology, development of digital teaching resources, modern educational technology, and multimedia courseware making.

HONGKUN ZUO was born in 1976. He received the master's degree from Anhui University. He is currently pursuing the Ph.D. degree with the Nanjing University of Aeronautics and Astronautics. He is an Associate Professor with the School of Finance and Mathematics, Huainan Normal University. His research interests include teaching theory and methodology of mathematical modeling, as well as financial quantitative investment analysis and financial engineering practice teaching, and numerical solutions of nonlinear dynamic systems and differential equations. He has guided students to win one second prize in the National Undergraduate Mathematical Modeling Competition, three second prizes in the competition area, and two third prizes in the competition area.

FEI PENG was born in 1976. He received the master's degree from East China Normal University. He is currently an Associate Professor and the Vice President of the Computer School, Huainan Normal University. He has published many articles. He has been outstanding in instructing college students in information technology skills competitions. He presided over a number of provincial and ministerial teaching and scientific research projects, and his courses include graphic and image processing, web design, and UI design. His research interest includes computer educational applications. He won the Teaching Achievement Award of Anhui Province for many times.