ADVANCING INFORMATION LITERACY IN HIGHER EDUCATION :A PREDICTIVE APPROACH TO ANALYSIS LEARNING BEHAVIOUR

[1] S.R. Prakalya, [2]S.Primathi

[1,2] Panimalar Engineering College, Chennai.
praks2203@gmail.com, primathi29@gmial.com

ABSTRACT:

Information literacy is an essential skill for self-learning and lifelong learning, and it is a fundamental ability for college students to adjust to the social demands of today. Using the rich and varied information literacy learning behavior features to do the learning impact prediction analysis is a useful method of exposing the information literacy teaching mechanism. By building a predictive model of the learning impact based on information literacy learning behavior characteristics, this article examines the features of college students' learning behaviors and investigates the predicted learning effect. Data on information literacy learning from 320 college students from a Chinese university was used in the trial. The qualities of information thinking and learning effect are significantly correlated, according to an analysis of college students' information literacy learning behavior using the Pearson algorithm. To categorize and forecast the learning impact of college students' information literacy, supervised classification algorithms such Decision Tree, KNN, Naive Bayes, Neural Net, and Random Forest are employed. In terms of learning effect classification prediction, the Random Forest prediction model is found to perform the best.92.50% accuracy, 84.56% precision, 94.81% recall, 89.39% F1-score, and 0.859 Kapaa coefficient are the values obtained. In order to improve the quality of information literacy instruction, optimize educational decision-making, and support the long-term development of exceptional and creative talent in the information society, this paper proposes differentiated intervention suggestions and management decision-making references for college students' information literacy instruction.The results of our research on the direction and way of thinking behind the sustainable development of information literacy training were promising.

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

INTRODUCTION:

The 21st century requires mastery of communication, network technology, and computer technology due to the rapid development of information technology [1]. Information literacy is a crucial component of college students' core literacy in the information age. A sort of adaption to the information society is information literacy. The cultivation of inventive skills and the sustained development of future talents are closely linked to college students' information literacy [2], [3].Cultural literacy and general quality include information literacy. Developing the information literacy of college students has already emerged as a significant problem in contemporary higher education. Along with information awareness and social ethics, information literacy encompasses the fundamental understanding and abilities of information and information technology, as well as the capacity to use it for learning, collaboration, communication, and problem-solving. People from many areas of life are currently interested in information literacy education. To varying degrees, information literacy instruction has been implemented by education departments and libraries in the US, UK, Australia, and other nations.Together with four other Chinese departments, the Ministry of Education released the "key points of improving the digital literacy and skills of the whole people in 2022" in 2022. In the upcoming years, it is anticipated that students' digital and information literacy will continue to advance. [4].Information literacy has also drawn increasing amounts of academic interest in recent years as a result of the growth of artificial intelligence technology, the impact of online and hybrid learning, and other factors. In order to implement focused information literacy instruction, numerous colleges and institutions both domestically and internationally have launched information literacy courses in a variety of ways. Sun Yat-sen University's "Information Literacy General Course - A Compulsory Course for Digital Survival," Tsinghua University's "Information Literacy: A Compulsory Course for Academic Research," Wuhan University's "Information Literacy and Practice - A Pair of Academic Eyes," and Sichuan Normal University's "Information Literacy and Lifelong Learning (Autonomous Mode)" are all available on the University of China's MOOC platform. [5] Given how college students are currently taught information literacy, a number of issues have surfaced.

Learning prediction is a highly relevant topic in the big data sector of education. One of the fundamental problems in learning analysis is the prediction of learning effects. Its main idea is to forecast the learning effect using a variety of data produced by students during the learning process and the machine learning approach. The prediction findings show that teachers can quickly ascertain the learning status of their students and make necessary interventions in the learning process. For instance, modifying teaching methods, enhancing students' study habits, etc. Hao and Wufati.[6] From principle research and application value to application in learning behavior analysis, data visualization, and learning prediction, learning analysis technology has advanced (Hang et al. [7].

According to AlShammari et al., learning prediction is predicated on learning achievement, learning objectives, and learning capacity. It also forecasts learning experience and learning impact based on the traits of learning behavior prior to and following learning. [8]. Theoretical models, empirical studies of prediction models, algorithm comparisons, algorithm development, early warning factor research, literature reviews, and more are all included in the prediction of learning outcomes.

Regression analysis, neural networks, Bayes, and other techniques are used to predict students' learning outcomes and performance. Gangshan and Gaihua [9]. Artificial intelligence technology can assist school institutions in using data to promote equity and quality in education, according to UNESCO's 2019 report, Artificial Intelligence in school: Challenges and Opportunities for Sustainable Development [10]. Using machine learning and educational data mining technologies to create learning impact prediction models in a

TABLE 1. Abbreviation of professional terms.

data-driven manner—that is, by automatically learning from data to create prediction models—is the current research trend and area of attention.

Based on how college students learn in information literacy classes, this study combines several distinct behavioral data points to produce an integrated data link. College students' learning effects are categorized and predicted through predictive analysis and evaluation of several machine learning classification models. The following inquiries are the main topic of this investigation.

(1) What are the better predictors of learning effects among college students' information literacy learning behavioral characteristics?

(2) Which machine learning models, according to the study sample, perform better and are more effective at making predictions?

(3) In addition to the study findings, what diagnostic observations were made for use in instructional interventions and learning recommendations?

II. LITERATURE REVIEW:

In order to increase college students' information literacy, this study reviewed the literature on the analysis of learning behavior features and the prediction of learning effects.

The literature data is mostly based on pertinent research conducted in the last three years and is primarily sourced from common databases for international paper retrieval, such as Web of Science, Scopus, Ei Compendex, etc. As indicated in Table 1, the reference texts are uniformly described due to the abundance of professional terminology and machine learning algorithm terms. This study examined the literature on the analysis of learning behavior aspects and the prediction of learning effects in order to improve the information literacy of college students.

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

The majority of the literature data comes from common databases used for worldwide paper retrieval, including Web of Science, Scopus, Ei Compendex, and others, and is based on relevant research that has been done in the recent three years.

The reference texts are universally characterized because of the prevalence of machine learning algorithm phrases and professional vocabulary, as shown in Table 1.

development of certain information literacy skills; online courses; or the connection between information literacy and intelligent environments. Research methodology literature comparison: While some researchers employ machine learning technologies, the majority mostly use quantitative and qualitative research, questionnaire surveys, data mining, and manufacturing quality experiences. The application of machine learning is indicated by the checkmarks in Table 2. The literature claims that researchers rarely employ machine learning techniques to conduct pertinent research and instead mostly employ traditional research methodologies.

Comparing the literature in order to find: Researchers have produced productive exploratory findings. To improve the use of information literacy, for instance, some researchers develop prediction models, some investigate novel approaches to teaching information literacy, and some design an appropriate teaching modality. Few advancements have been made in the analysis of information literacy learning behavior and the development of learning effect prediction models, according to the literature. Table 2 provides a summary of the comprehensive comparative research.

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

Researchers have studied the learning effect and learning behavior analysis from a variety of perspectives. The following are specific literary analyses and comparisons:

Comparing the literature in the field of study: The study primarily focuses on curriculum evaluation, learning quality analysis, teaching model effect, learning performance, and teaching effect prediction model research.

Literature comparison on research methodology: Neural networks, decision trees, support vector machines, and other machine learning algorithms make up the majority of research methodologies. Table 3's checkmarks indicate the application of machine learning. Individual investigations employed quantitative research and thorough analysis.

Literature comparison on the following findings: Numerous study findings are based on machine learning algorithms. For instance, models for predicting performance, analyzing students' willingness, predicting the impact of classroom instruction, diagnosing learning behavior, etc.

Researchers have studied learning behavior analysis and the learning impact from a variety of perspectives. The following are specific comparisons and analyses of the literature:

Comparing the literature in the field of study: The study primarily focuses on curriculum evaluation, learning quality analysis, teaching model effect, learning performance, and teaching effect prediction model research.

Comparison of research methodology literature: The majority of research techniques rely on machine learning algorithms, such as support vector machines, neural networks, and decision trees; Individual investigations employed quantitative research and thorough analysis.

Literature comparison on the following findings: Numerous study findings are based on machine learning algorithms. For instance, models for predicting performance, analyzing students' willingness, predicting the impact of classroom instruction, diagnosing learning behavior, etc.

In conclusion, theoretical deduction and experience form the basic basis of modern information literacy research, which establishes the hypothesis that certain elements are associated to academic success. Data is then gathered through questionnaires and interviews in order to examine 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

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

validate the theory. It is challenging to ascertain the quantifiable relationship between certain elements and academic accomplishment, and this approach can only demonstrate the correlation between these factors and academic achievement. There is a dearth of data intelligence analytic research on information literacy education, and machine learning and data mining technologies are rarely applied. To create models for predicting academic accomplishment, some researchers employ decision trees, neural networks, and other algorithms; nevertheless, there is a dearth of research on the impact of learning literacy on prediction.

Learning prediction research has great technical support thanks to the ongoing development and maturation of intelligent technologies like data mining, emotion analysis, and pattern recognition, particularly when combined with machine learning technology and the field of education. The use of educational data mining and other technologies is still the current research trend, despite several studies pointing out the detrimental effects of artificial intelligence on educational research.

Building an information literacy learning behavior characteristic analysis and learning effect prediction model for college students that is easy to use, has good prediction performance, and allows for differential recommendation and intervention based on prediction results is therefore an urgent problem.

III. MATERIALS AND METHODS

A. RESEARCH TOOLS

Among the often used learning prediction technologies are Rapidminer, Python, SPSS, Weka, and others. Model performance evaluation, feature set selection and classification prediction, and data preprocessing are the primary uses of SPSS and Rapidminer analysis tools in this work. Machine learning is the primary application for Rapidminer. The most popular data mining and machine learning program in the world is called Rapidminer. It has robust machine learning algorithms and offers features including data preparation and visualization, statistical modeling and predictive analysis, assessment, and deployment[39].

B. RESEARCH GOAL AND SOURCE OF DATA

The requirements for information literacy vary by location due to regional characteristics. As a result, general criteria should not be the only ones used when creating information literacy standards [40]. In the earlier study, the research team developed an index system for evaluating the information literacy of college students [41]. For this paper, the index system offers a fundamental reference tool. The research team developed the information literacy learning behavior characteristics observation scale for college students based on this assessment index after observing, measuring, extracting, and characterizing the information literacy learning behavior characteristics of college students. Awareness and attitude, knowledge and skills, application and creativity, ethics and accountability are all included in the scale. The comprehension of the significance of information technology is the primary focus of awareness and attitude. Information technology knowledge and skills are the primary emphasis of knowledge and skills. The primary focus of application and innovation is on creative use of information technology and cognitive thinking.

Information laws, rules, and moral principles are the primary emphasis of morality and responsibility. There are 28 indicators at the third level, nine indicators at the second level, and four indicators at the first level. This study categorizes students' learning scores into five groups: exceptional (5), good (4), medium (3), qualified (2), and unqualified (1) in order to assess the learning impact of the students. Likert's five-level scale is matched to each of the three-level indicators of information literacy learning behavior traits of college students: "1=never," "2=seldom," "3=sometimes," "4=often," and "5=always."

The observed markers of college students' information literacy learning behavioral traits are detailed in Table 4. Together with the aforementioned, the four domains of learning behavior are explained as follows: learning behavior in knowledge and skills, learning behavior in application and invention, learning behavior in morality and responsibility, and learning behavior in consciousness and attitude.

Information perception consciousness (IPC), information application consciousness (IAC), and lifelong learning consciousness (LLC) are the main components of learning behavior in consciousness and attitude.

Identifying and categorizing information (IPC1), finding, filtering, and evaluating material online (IPC2), and assessing the accuracy and dependability of information sources (IPC3) are examples of specific actions. applying knowledge and techniques connected to information technology to solve problems (IAC1); using tools like mind mapping to support learning (IAC2); Using Technology to Promote Professional and Personal Development (LLC2); Leveraging Technology to Promote Lifelong Learning (LLC1).

Learning behavior in terms of skills and knowledge: primarily including information application skills (IAS) and information science knowledge (ISK). Understanding various operating systems, word processing software, graphics and image processing software, and video and audio processing software operation methods are examples of specific behaviors (ISK1); comprehend the evolution, current state, and potential trends of information technology (ISK2); Understand the fundamentals of information retrieval and assessment, classification, and storage techniques (ISK3); Learn the fundamentals of data literacy, visual literacy, information literacy, and other multiliteracy (ISK4); To locate the necessary information, use a variety of search engines and network platforms (IAS1); Sort the data into categories and display it in tabular style (IAS2); information identification and analysis using a variety of techniques and procedures (IAS3); Make useful informational resources centered on certain educational subjects or based on particular teaching material (IAS4).

Applying and innovating behavior learning: primarily of information behavior (IB) and information thinking (IT). Define and recognize implicit assumptions in information, as well as infer information (IT1), are examples of specific actions. Implement successful instructional activities and carry out targeted information-based instructional design (IT2); leveraging information technology to assist with management and services (IT3); Create answers to problems by utilizing appropriate algorithms and integrating resources (IT4); Create and manage material using collaborative technologies like shared documents, project management software, etc. (IB1);Utilize cutting-edge means for communication (such as video conferencing, data sharing, and application sharing) to interact with others (IB2); create creative educational apps (IB3); to conduct cooperation and exchange in information technology (IB4).

Acquiring moral and responsible behavior: primarily encompassing information laws and regulations (ILR) and information ethics (IE). Among the specific behaviors are: Using learning resources in a healthy and appropriate manner to establish a positive information learning environment (IE1); controlling one's own information-ethical behavior and monitoring those of others (IE2); Respect the network civilization convention, clean up the network language, and engage in civilized and courteous communication and learning (IE3); impart knowledge of the rules, laws, and ethics pertaining to the use of technology (ILR1); Make it clear that everyone has equal access to information and that the intellectual property rights of others are respected (ILR2).

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)

C. METHOD OF RESEARCH

This study primarily consists of data preprocessing, feature extraction, method selection, model training, performance evaluation, and result analysis in accordance with the general process of learning analysis and machine learning.

Figure 1 depicts this study's primary technical path:

(1) Based on data cleaning, the correlation between learning behavior characteristics and learning effect is computed; observe and examine the relationship between learning effect and predictive factors, and determine the feature subset involved in model creation.

(2) The ten-fold cross-validation approach is used to train and evaluate the five models.

(3) By repeatedly tweaking the algorithm parameters, the prediction effect of a single prediction model can be enhanced.

(4) Determine the best prediction algorithm model by conducting comparative analysis and evaluation of prediction effects.

D. Gathering and preprocessing data

1) Information Gathering

In 2022, the study group of the "Research on Information Literacy of College Students Supported by Smart Campus," a teaching quality initiative in Anhui Province, China, conducted a "Special Survey on Information Literacy of College Students," which provided the research data. The study, which includes students from a range of academic and professional backgrounds, investigates the effects of dispersed, random, and representative data on the student population. In 2020, Huainan Normal University's 320 junior students' information literacy course performance data and learning behavior questionnaire data were used to get the data. A web-based questionnaire was used to gather data, and students in each class were given it in batches.

The questionnaire's overall recall quality was very high since a pre-survey was carried out before it was disseminated to determine whether the subjects properly understood the questions, whether the expression was suitable, and how cooperative they were. The data exhibits a positive distribution, minimal variation, and strong validity and dependability.

2) PREPROCESSING DATA

A descriptive statistical overview is presented in Figure 2. Numerical numbers are represented by the vertical axis, whereas variables are represented by the horizontal axis. It provides some indication of each variable's data outcomes. A few outliers and missing values were found via the descriptive statistics. The figure displays each feature subset's Min, Max, Average, and Deviation. For IAC1, IPC3, and IPC2, Average ranks in the top three, while IB4, IB2, and IB3 rank in the bottom three; for Deviation, IE1, IPC1, and IAC1 rank in the bottom three, while IB4, ISK2, and ISK3 rank in the top three.

Data preparation, which included operations like missing value processing, aberrant data processing, and data transformation, was done on the collected learning behavior characteristics and performance data to guarantee the caliber of the categorization learning model design. processing outliers. The training set's null data and other anomalous data were eliminated using SPSS.

The box plot is then used to eliminate the data's outliers. 315 captured data were ultimately kept after data cleaning.

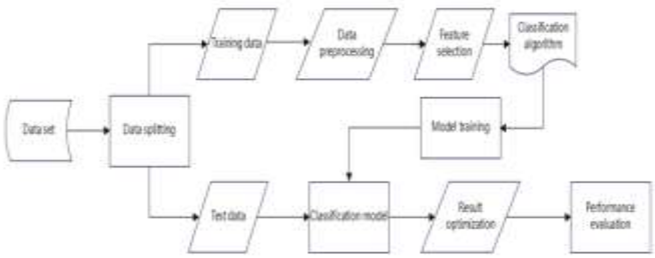


FIGURE 1. Technical road map.

transformation of data. Operationally transforming the data in the obtained training set is necessary to allow the machine learning model to achieve better recognition. The term "nominal attribute" should be used to describe the learning impact attribute. The qualities at the transformation level are "5=Excellent," "4=Good," "3=Medium," "2=Pass," and "1=Fail."

IV. RESULTS AND DISCUSSION

A. correlation analysis of the characteristics of learning behavior and the learning effect

Correlation analysis of learning behavior traits and learning impact can be used to model feature subset selection. The study of two or more components of related variables as a gauge of their level of relationship is known as correlation analysis. For correlation analysis to be carried out, related items must have some sort of association or likelihood. We can claim that two variables have a high correlation if they are strongly interdependent. Both groups are considered to be positively correlated if their values rise simultaneously; a negative correlation occurs when one group's value rises while the other group's value falls.

The correlation is computed here using Pearson's technique. With a correlation between -1 and 1, Pearson's correlation coefficient is a crucial indicator of how two variables are related to one another. The number of correlation coefficients acquired is as follows if there are P linked variables and the correlation coefficient between the two variables needs to be determined:

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

The correlation matrix can be found by arranging the variables in numerical squares according to their numbering. The correlation coefficient above the diagonal has a symmetric relationship with the section

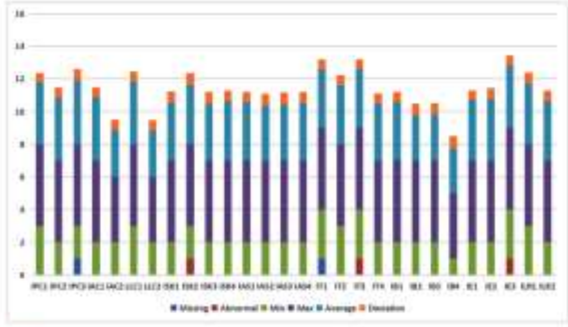


FIGURE 2: Dataset descriptive statistics

below, and the two identical variables on the diagonal from top left to bottom right both have values of 1.

The linear correlation between the current variables was measured by calculating the Pearson correlation coefficient between each variable and the learning effect. Figure 4 displays the correlation coefficients between the variables. The significance plot is the intersection of the two variables in the rows and columns, and the correlation coefficient is indicated by the color knob at the bottom. Figure 3 illustrates the relationship between the learning effect and the predictor factors. R accepts values in the range of -1 and +1. The two variables are said to be positively correlated if r>0, meaning that the larger the value of one, the larger the value of the other; conversely, if r<0, the two variables are said to be negatively correlated, meaning that the larger the value of one, the smaller the value of the other. Stronger correlations are those with bigger absolute values of r; weaker correlations are those with smaller absolute values of r [42].

By computing Pearson's correlation coefficients between the variables of college students' information literacy and the learning effect, the linear correlation between the variables that were already present was determined. The great majority of predictor factors were shown to have some positive link with learning outcomes. The examination of learning behavioral traits is supported by the correlations, which show some variety.

1) Examining the characteristics of high correlation learning behavior: IT1, IT4, IAS1, IT2, and IT3 are the top five factors. It is evident that the most important relationship between information thinking (IT) and the learning effect in application and innovation.

2) IT1 had the strongest association with the learning effect (0.810), followed by IT4 (0.768) and IAS1 (0.766). IT1 is the definition, detection, and inference of implicit assumptions in information. IT1 focuses on targeted critical cognition and information mining. IT4 is employing

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	0.383	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	0.383	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.

logical algorithms and resource integration to build problem solutions. The creation of information is the focus of IT4, which also emphasizes the growth of creativity and particular solutions to issues. To find the necessary information, ias1 uses a variety of search engines and online resources. It is a skill for applying information. The capacity of college students to obtain knowledge is the main emphasis of this learning behavior. The analysis's findings point to the improvement of knowledge and abilities related to information acquisition as well as the development of application and innovative skills. Including techniques for critical thinking in the information literacy curriculum Zhuozhuo et al. [43] is a crucial tactic to boost college students' capacity for creativity and invention and to overcome the information literacy education bottleneck. The development and enhancement of students' information literacy knowledge and abilities, particularly the enhancement of information acquisition skills in the information age, should continue to get emphasis.

3) EXAMINATION OF CHARACTERISTICS OF LOW CORRELATION LEARNING

ILR2 had a correlation of 0.430 with the learning effect, LLC2 had a correlation of 0.484, and IPC1 had a correlation of 0.486. LLC2 is employing information technology to enhance professional and personal development, whereas ILR2 is learning equal access to information and respecting others' intellectual property rights; Information is identified and categorized by IPC1. These three learning behaviors are abstract from the standpoint of learning behavior presentation space; from the standpoint of learning behavior presentation time, they are less integrated with the study, life, and current learning environment of college students than other learning behavior characteristics. This serves as a guide for future pedagogical development, which calls for college students to respond to these indicators with more efficient and successful learning practices.

4) DEFINITION OF A SUBSET OF CHARACTERISTICS IN LEARNING BEHAVIOR

Three learned behavioral features—IPC1 (0.486), LLC2 (0.484), and ILR2 (0.430)—with correlations less than 0.500 were excluded from the prediction model generation process in order to improve prediction models and produce better prediction outcomes with fewer features. Additionally, the gender variable was excluded from the prediction model construction due to its -0.103 degree of association with the learning effect.

In conclusion, a subset of 25 elements of college students' information literacy learning behavior are still in place, including IT1, IT4, IAS1, IT2, IT3, ILR1, IAS4, IB1, ISK4, IE1, IE3, IAC1, ISK3, IPC2, IAS3, IB2, LLC1, IE2, IAS2, IPC3, ISK1, IB3, ISK2, IB4, and IAC2.

B. LEARNING EFFECT PREDICTION USING CLASSIFICATION MODEL

The binary classification model's performance evaluation criteria include F1-Score (F1), Accuracy, Precision, and Recall. [44]. The numbers of positive samples with correctly predicted learning effects are denoted by TP, the number of negative samples with correctly predicted learning effects by TN, the number of samples that were mistakenly predicted as positive by FP, and the number of samples that were mistakenly predicted as negative by FN.

The percentage of correctly predicted samples in the classification model is known as classification accuracy, and it indicates how accurate the classification as a whole is.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

The precision, or the percentage of true positive instances among all results predicted as positive, is calculated by dividing the number of positive cases properly predicted by the classification model by the number of all positive cases projected by the model.

Precision = TP/(TP + FP)

The recall, or the percentage of true positive cases that the classification model detects, is the ratio of the number of positive samples that the model correctly predicted to the total number of positive samples in the test set.

Recall =TP/(TP + FN)

Precision and Recall are used to provide the F1-Score, a comprehensive metric. The F1-Score is a strong complete evaluation metric, and the higher the value of this metric, the better, because Precision and Recall are two measurements that contradict each other, and various problems focus on different criteria.

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

One indicator of classification accuracy is the kappa (KIA) coefficient. Kia is an index that makes it possible to compute categorization and overall consistency. An evaluation of a multi-classification model's accuracy is conducted using the KIA. The model achieves a higher classification accuracy the larger the value of this coefficient. The kappa coefficient can be computed in this way. Po stands for the percentage of consistency or accuracy cells in observations. The percentage of cells that are contingently consistent or anticipated to be contingently consistent is denoted by Pc.

$$KIA = (Po - Pc)/(1 - Pc)$$

Predicting college students' learning impact levels using their information literacy behavioral performance indicators is the primary goal of prediction model selection. Since this is a common classification problem, the prediction performance of various models is compared independently for this research sample using the traditional machine learning classification algorithm. Ten-fold cross validation is used to train and evaluate the five models in the following. In order to maximize the use of samples, the dataset is first split into ten parts, nine of which are rotated as training data and the remaining one is used as test data. The model is then trained by averaging the correct rate each time as the evaluation value of the algorithm accuracy.

1) THE DECISION TREE

Using recursive binary partitioning on the feature space, the greedy decision tree algorithm classifies instances according to features. The sample data is compared with the decision tree's feature nodes starting at the root node. The branches at the following level are chosen to continue the comparison depending on the judgment result, and the classification result is the last leaf node [45]. Decision trees have the advantage of being quicker to classify and easier to read [46]. The "gain ratio" is used by the C4.5 decision tree technique to choose the best partitioning attribute.

When the core parameter Maximum depth is set to 8, the minimum leaf size is set to 2, and the confidence level is set to 0.1, the model's best recognition impact is achieved by training and optimization of the Decision Tree parameters. Figure 5 illustrates the experimental procedure for obtaining a decision tree. Examine the outcomes of several tests to obtain the ideal accuracy rate of 84.17%.

2) K-NEAREST NEIGHBOR

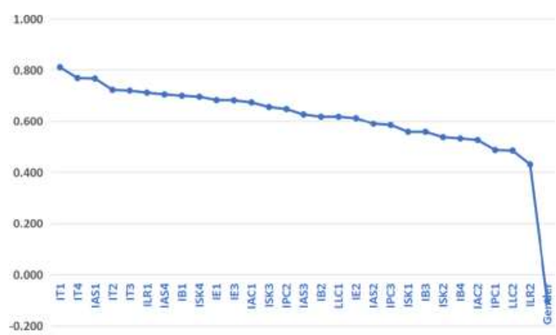


FIGURE 4. Correlation between predictor variables and learning effect.

A statistical classification-based approach is the K-Nearest Neighbor (KNN) algorithm. This algorithm's drawback is that it is insensitive to outliers, but it has the advantage of not having to divide the vector space made up of all data records and improving classification by using the model data to identify K similar vectors [47].

The model recognition impact is optimal when the core parameter K is set to 6 after training and optimizing the KNN parameters. Figure 6 displays the results of the KNN experimental technique. The results of several trials are examined in order to determine the ideal accuracy rate, which is 90.83%.

3) NAIVE BAYES

The Naive Bayes algorithm uses probability theory to detect and classify data. In addition to using sample data with previous information to calculate the posterior probability of occurrences, the technique is able to connect the prior and posterior probabilities of events. One of its benefits is that the model is easy to build and has excellent stability and efficiency [48].

The model recognition impact is optimal when the core parameter minimum bandwidth is set to 0.2 after training and fine-tuning the Naive Bayes parameters. Figure 7 illustrates the experimental process for getting Naive Bayes. The results of several trials are examined in order to determine the ideal accuracy rate of 90.00%.

4) The neural network

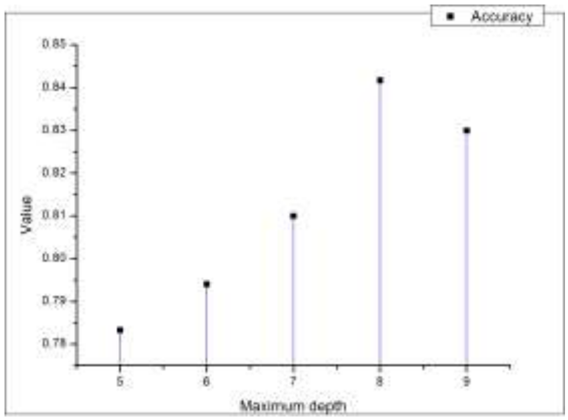
A neural network is a mathematical representation of a biological brain network used for information processing; neural networks are successfully used to solve classification difficulties [49]. The input layer, hidden layer, and output layer make up the majority of neural networks.

When the Neural Net parameters are trained and optimized, the model's best recognition effect is achieved with momentum set to 0.9, training cycles set to 200, and learning rate set to 0.01. Figure 8 illustrates the experimental procedure used to create Neural Net. The ideal accuracy rate, based on the findings of several trials, is 91.67%.

5) RANDOM FOREST

In order to increase randomness, prevent overfitting, and integrate the outcomes of a single decision tree in accordance with the rules of bagging, Random Forest uses random sampling of data samples and features to train multiple tree classifiers. This approach avoids learning all samples and all features per tree [50]. In order to create K classification regression trees, the training sample data is sampled using put-back. Assuming that the feature space contains n features, m features are randomly chosen at each tree's nodes, requiring m < n. Each tree is allowed to grow as much as possible without pruning, forming a forest with multiple trees, and the classification results are based on the number of tree classifiers that vote. The Random Forest model's best recognition result was achieved by training and fine-tuning its parameters. The criterion was set at gain_ratio, and the number of trees parameter was set to 150. Following multiple executions, Figure 9 displays the Random Forest's experimental outcomes. Examine the outcomes of several tests to achieve the best accuracy rate of 92.50%. Following each model's parameter adjustment, Table 5 displays the prediction results for each model. Figure 10 provides a visual representation of the prediction models. The kappa taking value range denotes varying degrees. 0.4~0.6: moderate; 0.6~0.8: significant; 0.8~1.0: nearly perfect; 0.1~0.2: minor; 0.2~0.4: fair [51]. Each model's overall consistency and categorization consistency are typical and essentially meet the requirements..

Figure 5. Selection process of decision tree



According to the examination of the prediction results, Random Forest has the highest Accuracy (92.50%), followed by Neural Net and KNN; Naive Bayes has the highest Precision (93.06%), followed by Random Forest and KNN; Random has the highest Recall (94.81%), followed by KNN and Decision Tree; Random Forest has the greatest F1-Score (89.39%), followed by KNN and Naive Bayes; Random Forest has the highest Kapaa (0.859), followed by Neural Net. According to the results of every indicator, the Random Forest prediction model performs the best and may be applied to improve the information literacy learning effect prediction of college students.

Figure 6:The selection process of KNN hyperparameters.

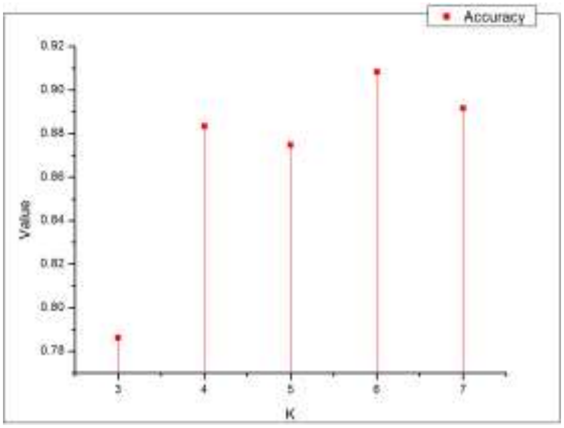


Figure 7: The selection process of Naive Bayes hyperparameters.

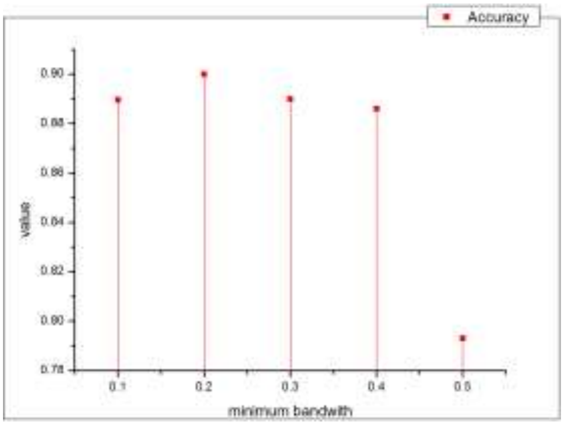


figure.8. selection process of Neural Net hyperparameters

.

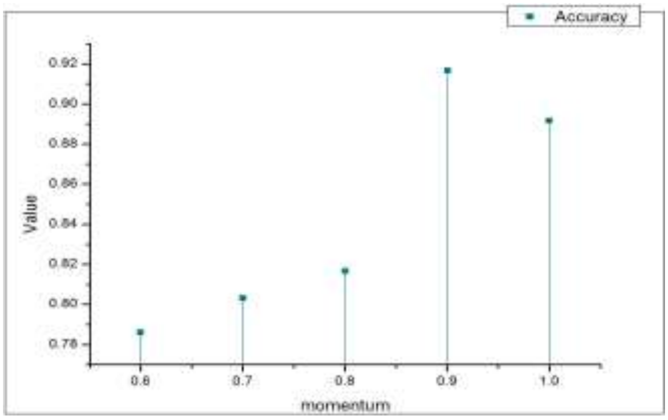


Figure 9. The selection process of Random Forest hyperparameters

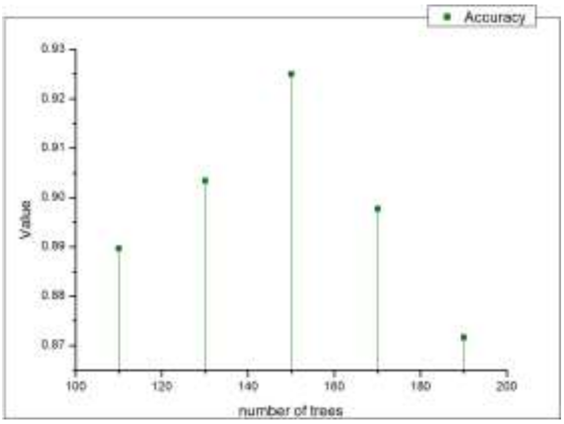
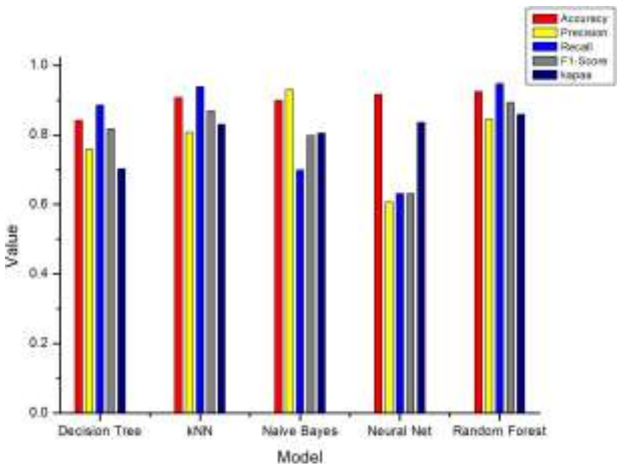


Figure.10. . Performance comparison of learning effect prediction models.



DISCUSSION

Exploring the qualities that can be explained and relied upon for improving the training process is more significant than only aiming for the high accuracy of the machine learning prediction model [52]. The correlation study clarifies the intervention strategies required to raise the standard of instruction for college students' information literacy. The learning behavior features of college students' information literacy are analyzed using the Pearson method, which shows that information thinking and information application skills—in particular, information thinking—have a stronger link with learning outcomes. The goal of teaching information literacy should be to foster students' creativity and practical skills. From the standpoint of information transfer, several colleges make an effort to foster critical thinking [53]. Some academics suggest integrating critical thinking into professional courses as part of an information literacy education model [54], adopting an information literacy education model that incorporates innovation and entrepreneurship training for college students to foster students' innovative thinking through the use of the "holistic thinking approach" [55], and enhancing critical reflection skills based on the "163" information literacy education system of "Internet+" [56]. Nevertheless, there isn't a comprehensive training program for critical and creative thinking skills.

The classifier produced by the Random Forest algorithm performs optimally in terms of prediction performance for all model types when the four metrics of accuracy, precision, recall, F1 value, and kappa are combined. In general, the results of comparable research like Juan et al. [57] and Faqin et al. [58] are consistent with the increased prediction accuracy, which demonstrates the impact of machine learning algorithms applied to learning effect prediction modeling. We can more precisely forecast the learning impact of college students' information literacy instruction, direct the modification of instructional strategies and resource allocation, and effectively ensure the quality of instruction by employing the Random Forest algorithm model for this purpose.

V. CONCLUSION:

The findings demonstrate that the prediction model put out in this work significantly influences college students' development of information literacy. On the one hand, a more significant association between information thinking, information application skills, and learning effect is revealed by an algorithmic analysis of the learning behavior features of college students' information literacy. The development of information thinking skills should be prioritized, but the development of information acquiring skills should not be overlooked. From the earliest stages of talent development and the cultivation of exceptional and creative talent, universities should recognize the significance and necessity of teaching college students information literacy. In order to provide intelligent learning tools and learning environments that support independent, cooperative, and research learning, integrate critical thinking techniques into the information literacy education system, create a long-term assessment mechanism focused on the development of critical thinking, and actively support college students' critical thinking cognition and knowledge creation skills, it is imperative that network and multimedia technologies be fully utilized. On the other hand, it is essential to streamline the learning materials.

Prediction results of the classification model.

TABLE 5.

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

and consistently instill and mentor college students to develop the idea of lifelong learning. In conclusion, this study suggests an effective machine learning approach to characterize the learning behaviors and predict the learning effects of information literacy among college students. It employs data-driven thinking to encourage 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. This study gives educational administrators a more reliable data base to analyze the potential connections between information literacy education phenomena and outcomes, ultimately increasing the success rate of educational decisions.

There are still restrictions even though this study has done some research. Although machine learning techniques are somewhat effective, the study contends that technological instruments must be used in accordance with the particular teaching and learning context. At a later time, more research is suggested in the following areas. (1) The quality of the learning behavior trait scale is under more pressure since learning impact prediction has not been able to account for other potential aspects in learning scenarios. More scenarios will be used to investigate the learning behavior characteristics, the learning behavior characteristics scale will be developed, the universality will be enhanced, and a closed loop between teaching experiment, teaching research, and teaching practice will be established. (2) Only five supervised classification algorithms— Decision Tree, KNN, Naive Bayes, Neural Net, and Random Forest—are employed in this study.

The implemented methods can be combined in later research to produce better prediction outcomes.

ACKNOWLEDGMENT:

The author would like to express gratitude to everyone who worked hard to make this research paper better. Additionally, they would like to express their gratitude to the anonymous editors and reviewers who helped make this paper possible.

REFERENCES

[1] Z. Changhai, ''Research on the information literacy education model of Chinese college students based on critical thinking and creativ- ity,'' *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. The- ory Pract.*, vol. 30, no. 10, pp. 38– 39, Oct. 2008.

[3] D.YanandG.Li,''Aheterogeneitystudyonth eeffectofdigitaleducation 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 experi- ence of the 'epidemic' of the education war—Online education perspec- tive,'' *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,''Miningthecharacteristicsofmea ningfullearningbehavior: 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 occur- rence process, design model and mechanism explanation of deep learn- ing,'' *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,andZ.Al-Shammari,''Ameta-analysisof 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 behav- ior 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 enlight- enment 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,''Machinelearning-basedevaluationofinformationliteracyenhance- ment 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,andS.Zhu,''Smartclassroom preferencesand 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,''Researchoncollegestudents'informationlit eracybasedonbig data,'' *Cluster Comput.*, vol. 22, no. S2, pp. 3463–3470, Mar. 2019, doi: 10.1007/s10586-018-2193-0.

[15] Z.Mian,Y.Bai,andR.K.Ur,''Researchoncollegecom puter-computing and information literacy online course based on MOOC: Taking the north minzu university as an example,'' in *Proc. IEEE 3rd Int. Conf. Com- put. 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.NishikawaandG.Izuta,''Theinformationtechnologyliteracylevelof 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,andD.Fujiani,''Theinfluence 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 infor- mation 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 under- graduate music therapy education: A cohort study,'' *Music Therapy Per- spect.*, 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 stu- dents' 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.