

Research on College English Teaching Model Based on Decision Trees

Hao Wu^{1,*} and B. Nagaraj²

¹Hunan Radio and TV University, Changsha, 410004, China

²Rathinam Technical Campus, Coimbatore, 632014, India

*Corresponding Author: Hao Wu. Email: whwendy@163.com

Received: 06 February 2021; Accepted: 28 April 2021

Abstract: English teaching has always attracted much attention. However, the processes of its transmission and acquirement is often divided into two separate parts, which seriously hinders the effective implementation of its objectives. Teachers attach particular importance to the choice of the curriculum structure and teaching material. Students are busy comprehending the assignments their teachers deem important. Under such a scenario, the effective acquisition of knowledge and the development of sustainable comprehensive abilities are ignored. The random forest algorithm in machine learning applications could play important role improving on the current English teaching system. A random forest model is constructed using a decision trees selection method, which focuses on 19 attributes of the English teaching model. Results show, to begin with, that the indigenous teaching plan and environment fail to adapt to the pace of knowledge iteration in the era of big data. Moreover, interactions between teachers and students appear to be shallow with little constructive interaction, causing a decline in the relationship between teachers and students. Last, there is still no signs of any legitimate construction in terms of in-person English teaching, relativity attribute, *corpus* and platform. Therefore, this paper has proposed a new English teaching model to adapt to the current college English teaching environment. The experimental results show that the method is effective and feasible for the current college English teaching.

Keywords: College English teaching; machine learning; attribute model; decision trees

1 Summary of College English Teaching Model

In English teaching, teaching and learning are considered to be 2 inseparable and correlated processes. If it is split into 2 separate parts, the relationship between teachers and students will be weakened, the transmission of knowledge will become difficult, and internal circulation will become ineffective. As a consequence, students will become disinterested in learning, thus rendering the outcome of learning unsatisfactory.

The analysis of attribute prediction is a crucial link between teaching and learning in the reform of college English teaching. There are four primary reasons for improving this work. Firstly, it is difficult for students to develop new cognition because teachers stick to the obsolete patterns of thinking and teaching



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

environments that are incompatible with the current era of big data. Secondly, the channels for students to acquire knowledge are not as smooth as required. Despite some modern facilities in the teaching environment, there are still few reasonable and effective applications. Thirdly, it is rare for teachers to prioritize teaching interaction and effective teaching evaluation systems, which disconnects teachers from students. This is because teachers are unable to provide students with accurate information in the first place, and students struggle to understand the context fast. Finally, there remain some problems with the influencing factors, the division of teaching entities, the definition of attributes relevant to teaching entities, information-based corpora, and teaching orientation. Therefore, it is imperative to construct a teaching model applicable to the college English teaching.

1.1 Teaching Entity and Attribute

In college English teaching, students are treated as the main body. Students develop through the reasonable integration of the teaching environment and correct guidance from teachers. In the teaching process, the main research object is the teaching entity, which includes students, teachers, and the teaching environment.

The implicit attributes of teachers include the teaching content, teaching skills, teaching materials, language literacy, theoretical literacy, teacher emotion, teaching skills, cultural awareness, cognitive strategies, and individual emotions. Meanwhile, the implicit attributes of the teaching environment include the teaching lever, the state of the curriculum, the state of the teaching management, teaching structure, objective requirements, organizational content, methods, task requirements, corpora, the use of tools, multimedia technology, and system evaluations. Lastly, the implicit attributes of students include their interests and hobbies, the ability to learn independently, the positive state of learning, driving force, beliefs, motivations, background knowledge, and the courage to think, along with the time and space of learning, social investigation, social practice, the ability to conduct scientific research, practice, and the degree of innovation.

1.2 Correlation Theories

The prediction and analysis model of attributes of the college English teaching process is based on the random forest theory, which is based on multiple decision trees. The decision tree algorithm itself has many characteristics, such as less training time and lower complexity, faster prediction process, and easier model display. In view of the overfitting problem faced by the single decision tree, the random forest has good adaptability.

1.3 Current Research Status

There have been remarkable achievements in research on the college English teaching model. Such research has focused on English teaching strategy, college English teacher training, the teaching method, communicative ability training, and teaching assisted by artificial intelligence. Through exploring the English teaching model, the cognitive ability of students and their capacity for innovation can be improved, so as to meet the needs of society. Besides, the overall quality of teachers and management institutions can be improved on a continued basis. Due to the profound impact of the English teaching model on the future professions to be chosen by students, many researchers around the world are paying much attention to this teaching model.

In Australia, the SFL (systemic-functional linguistics) framework was adopted by Stosic to study the difficulties encountered by Serbian language learners in developing practical English teaching strategies [1]. The data was obtained through English questionnaires and interview conducted with the teachers in Serbia. According to the research and analysis, learners will feel confused when tense forms reflect more than one tense. Thus, it is significant to determine how to improve teaching strategies for the learning of

a second language [2,3]. In Germany, it is requisite that teachers strengthen their own education and research on learning strategy. Through analyzing the significance attached to understanding the oriented learning strategy and the status of its application [4], this paper explores the characteristics exhibited by the learning of knowledge by students, such as error priority, the lack of coherence, and situational sensitivity. Then, the methods of knowledge education are proposed to tackle these problems, such as refutation strategies, comparative cases, and the assistance with classification. From the perspective of teachers, the knowledge related to other fields plays an equally important role in the education received by students. It is necessary for teachers to deliver targeted education based on the latest discovery of learning strategies and the knowledge acquired by students. In Iran, Zahra [5] takes the view that English has become an indispensable part of the curriculum implemented across Iranian schools since the 20th and 21st centuries. Though Iran has made efforts and investments, the results still fall short of expectations. Zahra conducted research on the problems arising from the English teaching process implemented in Iran, carried out an analysis of different potential factors, proposed a classification of these factors, and explored the solutions. In New Zealand, Hong et al. [6] carried out a unique research on LRP (language-related plots) of EST (English secondary teaching) by recording the in-class teaching activities. Their analysis showed that LRE (language-related episodes) is similarity in most cases. The leading course for LRE to occur frequently is the use of vocabulary under the guidance of teachers who actively make students pay attention to language learning, with teaching promoted through the constant development of content-based discussion. In Denmark, Dimova [7] conducted a study on the validity of scores for TOEFL and IELTS scores; as the aim was to explore the political, economic, socio-cultural, and academic dimensions of a specific ESE (English secondary education) environment, and to make practical arguments for standardized English tests. In Turkey, Petek et al. [8] conducted a survey by combining CT (critical thinking) with practical teaching. A 14-week action plan was explored. Eight English teachers were followed up on their understanding of the integration of CT into language teaching, and its impact on their teaching. Through the collection of data-regarding CT tasks, semi-structured interviews, curriculum plans, teaching videos, and self-reflection, it was found out that language teaching can be improved by adequate and appropriate guidance.

Over the course of an English teacher's career, their training is also important. Ismail [9] conducted a self-report questionnaire survey of 49 English interns in Turkey. As part of the training provided to English teachers, their study aimed at exploring the feelings of interns about their college education. The interns were very satisfied with their training plans, but they often complained that they had not experienced interaction with real students during their education. Therefore, a training plan can support teachers in coping with arising challenges, and listening to the concerns of teachers conducive to improving the practicality of future teacher training programs. There are significant differences between experienced teachers and new teachers in the process of teaching. Jansen et al. [10] conducted an analysis and evaluation of 36 experienced teachers and 47 junior teachers simultaneously; the judgments of these experienced teachers and junior teachers were compared against machine scoring and expert scoring. The judgment of experienced teachers was more negative than that of junior teachers, while other judgments were more negative than those of experts and machines. The result was consistent even when the knowledge content was controlled. The framework of professional knowledge acquired by teachers has a great impact on the outcome of teaching. In the context of higher education, the composition of professional knowledge acquired by teachers determines their future tasks and the direction of their development. Dijk et al. [11] analyzed and synthesized the professional knowledge framework adopted by 46 teachers through systematic investigation, and integrated the professional knowledge of university teachers with multi-dimensional development. It was found that was conducive to the coordination and development of teachers' professional knowledge and ability to perform different teaching tasks in the higher education setting.

The teaching method plays an important role in the process of English teaching. In order to expand the evaluative expression data of students, Unsworth et al. [12] developed a multi-modal creative teaching

method, involving a language framework applied to the way in which attitudes and meanings are expressed. Len believes that the language used by teachers and their attitude convey the knowledge of visual semiotics and the application of digital multi-modal creation software. The key point is that students enjoy developing their ability and confidence with creative tools, which can expand the multi-modal expression data on the attitude of students. In addition, the combination of facial expression and gesture is beneficial to English teaching. In order to establish the connection between linguistics and sociology, and contribute to the design of overall teaching methods and promote the critical orientation of the literature, Ryshina et al. [13] proposed the concept of SFL (system function linguistics). SFL is a tool used to the assist text analysis and the integration between textual content and language. Not only does it conceptualize the content of human and nature, it also supports learners in reflecting on cultural forms and applying language and knowledge in a creative way. Laying emphasis on the application of creative approaches to education in the ESP (English for specific purposes) teaching process, Rus [14] analyzed the importance attached to the change in roles and functions when attempting to innovate an entire education system, especially the learning of professional language and a series of methods in relation to innovation factors. Dana also explains the advantages of carrying out ESP teaching through the use of innovative methods. The application of creative teaching methods in ESP teaching plays a role in improving the student motivation. Moreover, it is a precondition for success in applying language skills under an ESP environment.

Student's comprehension is largely dependent on the improvement of communicative competence. To cultivate communicative competence, Ciornei et al. [15] aims at supporting students in improving their academic performance and gaining an understanding of the target culture not only through the establishment of relationships between the use of authentic texts and enhanced communicative competence, but also by determining the types of authentic texts. Imam [16] conducted a study on the identification of significant personality characteristics in English curriculum. Imam conducted 10 semi-structured interviews, which is purposed to cultivate the interest of students in specific curriculum contents, improve their English skills and make them to adaptive to higher education. Over the past few years, Kinya et al. [17] has conducted an investigation into the new education model called "Future Strategic Design" talent training. Its goal is to better clarify the enforcement of different strategic design education models. Various advanced technologies, including the Internet of things, artificial intelligence, augmented reality, and virtual reality, were applied to analyze not only the management scheme of teaching model design, but also the management program of business modeling with platform services, product development and strategy, as well as the customer service training. On this basis, a comparative study was conducted to determine the appropriate approaches to education. As globalization deepens, there are more and more adults from different countries seeking to learn from overseas English teachers, usually motivated by a desire for more effective communication at work, or to facilitate overseas travel and other social activities. Realizing these needs, Mihaela [18] collected a large amount of relevant information about English learners with a strong motivation to study, including the analysis of distinctions between their teaching models, the analysis of the comparative data on learners in different age groups, and the analysis of the characteristics shown by English learners in respect of cognition, attitudes, behaviors and methods. Their study is beneficial to those English teachers who need more knowledge about a foreign language. From a unified perspective, Vallente [19] analyzed the narrative and conducted interview with the research object, based on which the identity formation of pre-service English teachers in a multilingual environment was explored. The participants performed poorly in communication, while the performance of native English teachers was relatively standardized.

Using artificial intelligence to assist in English teaching is becoming more and more appropriate [20,21]. Zhou et al. [22] conducted a data analysis by combining machine learning with extreme learning machine. Due to the explosive growth of teaching resources networks, it is difficult for students to identify the valuable learning resources in a short time, and it is not easy for students to find the suitable reading materials.

Therefore, it is very important to find a solution to quickly obtaining the most needed learning materials. Arturo et al. [23] proposes a summary recommendation algorithm based on online multi-source documents to improve the text readability and produce a recommendation. The algorithm is designed to process a collection of documents related to web page query results. It has been tested in English and Spanish documents by using different terms and sentence correlation metrics. According to the experimental results, the algorithm removes the need for preliminary training, which makes it distinct from most traditional methods. Moreover, the experimental results show that the method is effective. Machine learning has also gained obvious advantages in numerical analysis [24,25]. Amid the Covid-19 pandemic, people have been exploring ways to estimate the number of cases. For instance, Yelkanat [26] studied the performance of a random forest machine learning algorithm in estimating the number of Covid-19 cases in 190 countries around the world, and mapped the actual number of confirmed cases. The number of confirmed cases from January 23 to June 17 in 2020 was divided into three sub-data sets: a training sub-data set, a test sub-data set, and a prediction sub-data set. The prediction model was proved to be feasible. Meanwhile, Tiwary et al. [27] proposed a method used to classify coal chemical composition and minerals at different stages based on the random forest. Firstly, the relevant features were extracted according to the macroscopic determination and grade measurement of optical properties. Then, the appropriate random forest model was established. Finally, a stable classifier was obtained through training and learning, which is a successful application of random forest algorithm.

In China, case teaching is widely practiced in English teaching. In an observational case-based teaching study, Li et al. [28] observed 16 hours of video of classroom teaching meetings, live recordings, and interviews. The data analysis revealed three key issues: the teaching process, the teaching conditions, and scientific elaboration. In order to cultivate students' ability to conduct scientific research, it is necessary to set up specialized English teaching. Zhang et al. [29] proposed a three-year ESP action research method. Constant reflection and revision allowed Zhang to enrich the method, which led to satisfactory results in the integration of corresponding curricula. In business English teaching, many teaching ideas have been put forward. However, little attention has been paid to the influence of teaching ideas. Through the research of pedagogy orientation [30], Chan et al. [30] developed a new teaching concept and made it consistent with the actual teaching situation, and analyzed its impact value. The research materials involved oral English recordings, business English courses in some universities, and the real data on students' role-playing. Liu et al. [31] proposed a method of non-destructive detection of sunset yellow in cream, which is based on infrared spectroscopy and interval random forest. The research results show showed that interval random forest can find the optimal sub-interval quickly and non-destructively, thus confirming the prediction ability of the model. As the tourism industry grows, the demand for English professionals is increasing. Communicative English teaching can promote English learning for students, improve their self-confidence, and enhance their communicative competence in tourism English [32]. Communicative English teaching can provide practical and effective guidance on practical application. In terms of oral English pronunciation, a metacognition strategy is the key, which includes self-reflection, control, evaluation, and adjustment. Li [33] explored these characteristics and in college English teaching based on metacognitive processes simulation, and proposed a mathematical model to improve pronunciation for students through quantitative and effective evaluation.

2 Random Forests in College English Teaching Model

2.1 Random Forests Decision Tree Theory

Random forest is a combined classifier algorithm proposed by Breiman in 2001. A classification and regression tree (CART) is used as a meta classifier. In order to solve some problems in the process of college English teaching, this paper focuses on finding the problems as soon as possible and reversing the

possible adverse effects at an early stage. Through using the powerful data analysis ability of the random forest, the present paper provides scientific methods for predicting the actual teaching effect.

2.2 Research on the Properties of College English Teaching Model

According to the entity objects of “curriculum teaching”, this paper explores the atlas structure model, which can accurately reflect the essential characteristics of the entity; thus, we can formalize the description and reasoning of the entity. We then transform the atlas structure into a tree model, and establishes a college English teaching attribute model matched with the real semantics.

2.3 Prediction about Abnormal Problems of Random Forests

The classification of machine learning algorithm [34] is used to establish the prediction model, which involves support vector machines, decision trees, artificial neural networks [35], and self-enhancement methods. Random forest is an integrated learning algorithm with good classification performance, higher classification accuracy, and high computational efficiency; therefore, it is suitable for the calculation of various data sets. The random forest also has good robustness to feature selection of entity attributes in college English teaching, good performance in high-dimensional feature vector space, strong data reasoning, and normalization ability. Hence, it is feasible to use the random forest as a prediction method for abnormal problems in college English teaching.

This paper mainly focuses on the prediction of abnormal learning behavior in the process of college English teaching. Firstly, it studies the structure of the attribute model centered on each entity object, and obtains the simplified and effective feature data through data, preprocessing, reduction, and other operations. Then, the rule attribute analysis is carried out to obtain the feature data; the similarity measurement of each entity attribute is used at the same time, and the allocation of attribute weight is considered. Finally, the optimized teaching model is established based on the hierarchical prediction model. The overall technology roadmap is described in Fig. 1.

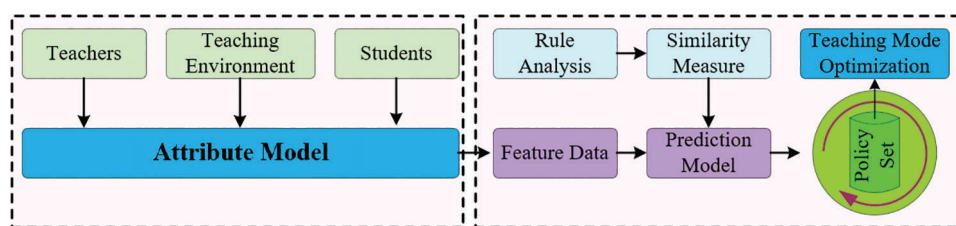


Figure 1: Overall technology roadmap

2.4 Classifier Learning

In the classifier design, the prediction accuracy of a single classifier is not high, overfitting appears easily, and the generalization ability of the trained classifier is weak. Through the combination of multiple classifiers, the performance of the whole classifier can be effectively improved. The independent classifiers need to meet two conditions: the accuracy of the independent classifiers should be better than the random guess; and the independent classifiers should be different from each other. Different training sample subsets can be repeatedly obtained from the original sample set, and can train each independent classifier. If the size of the original sample set is N , then the probability that each sample is not extracted as $((N-1)/N)^N$. When N is large enough, its value approaches to 37%, the accuracy of the classifier is improved.

2.5 Random Forest Algorithm

The random forest is based on the K decision tree $\{h(X, \theta_k), k = 1, 2, \dots, K\}$ as the basic classifier, and the combined classifier is obtained after classifier learning. When the input samples are classified, the random forest classification results are output for each classification through voting of the simple decision tree. Where $\{\theta_k, k = 1, 2, \dots, K\}$ is a sequence of random variables, which is composed of a random thought and random forest decision.

The training process is the essence of random forests for training each tree, and the training of decision tree is independent to each other; thus, the random forest can realize its training through parallel processing to improve the efficiency of the model. The training process of the k -th $h(X, \theta_k)$ random forest tree is shown in Fig. 2.

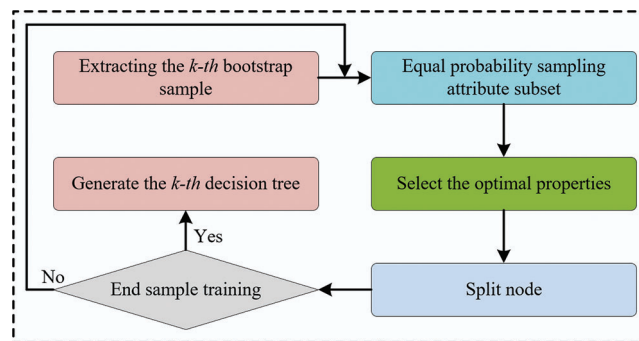


Figure 2: Random forest single tree training process

3 Research on College English Teaching Model

3.1 Analysis of College English Teaching Model

The model includes subject information, course learning management, and teacher information. Subject information is mainly composed of student information, curriculum information and teacher information. Course learning management is mainly composed of three parts: students' classroom performance, students' work performance, and students' academic performance. Teacher information is mainly composed of teachers' experience information, teachers' observation information, and teachers' intuitive information. The instruction method of the teaching model is shown in Fig. 3.

A questionnaire of 19 items is compiled. The items can be divided into three: structured data of research subject information, semi-structured data of research course learning management, and unstructured data of teaching information. Details on the questionnaire are given below.

(1) Investigation purpose. According to the quantitative rules designed in this study, and combined with the machine learning method of RF prediction, the paper provides the guidance for the college English teaching model. The random forest structure is shown in Fig. 4.

(2) Respondents. A total of 240 English majors from graduation years 2017–2019 were selected as the subjects of the questionnaire.

(3) Investigation content. Content corresponded to 3 key areas: subject information, course learning management, and teaching information. The independent variables in the survey design were all based on a Richter scale 5 (1 = very poor, 2 = poor, 3 = normal, 4 = good, 5 = very good), and the dependent variable was rated 2. These 19 variables were involved in the college English teaching model; some of these variables are shown in Tab. 1. Note that the variable assignment was reduced from the highest, the most and the best, and * was a 5-level subjective scale without quantitative standards.

(4) Effective data collection. The number of distributions was 3 times, 240 copies each time. From 2018 to 2020, a total of 720 questionnaires were issued, and 682 valid questionnaires were returned. The prediction rules were obtained according to the scores in the survey contents.

(5) Data preprocessing and analysis. Data preprocessing can effectively select noisy, inconsistent, and incomplete data, and includes four basic steps: selection, consistency, cleaning, and discretization. The first three steps solved the superficial problems existing in the data, and the fourth step involved the connotation of data. The methods of discretization included cleaning, integration, conversion, and reduction. The split attribute of each decision tree in RF was predicted according to the Gini index value.

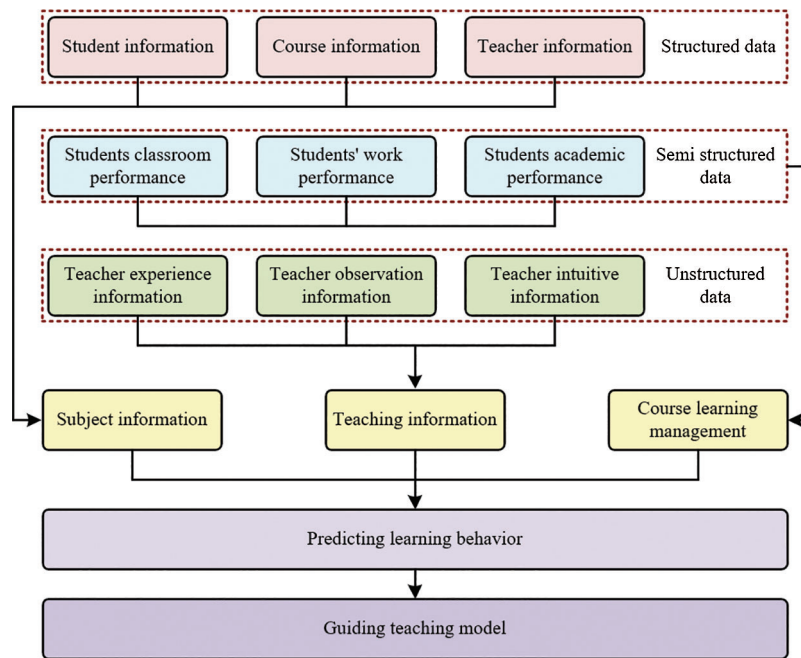


Figure 3: Instruction method of the teaching model

3.2 Experimental Results Analysis

We obtained the general trend of characteristic data from analyzing the survey data. The analysis results are shown in Fig. 5.

The relationships among features are also important factors. In the matrix diagram shown in Fig. 6, we can observe that the learning time has a strong proportional relationship with the actual results.

The top 14 features are shown in Fig. 7 according to their importance.

These 19 attribute characteristics (M) are set in the questionnaire. According to the requirements of random forest row and column sampling, the number of m features selected in the attribute feature column is about $SQRT(M)$, where M is the total number of features, and $SQRT()$ is the root mean square function. When $m = 2$, the recognition accuracy is the lowest, which is 93.4%; when $m = 6$, the recognition rate is 95%. After more than 6 times, the accuracy slightly fluctuates at about 95%. Therefore, $m = 6$ is suitable for this experiment. Fig. 8 shows the change in the number of classification subtrees and reasoning accuracy under the value of different characteristic number m . When the number of subtrees reaches a certain scale, the number of subtrees has little influence on the results. In this experiment, the number of subtrees is 90.

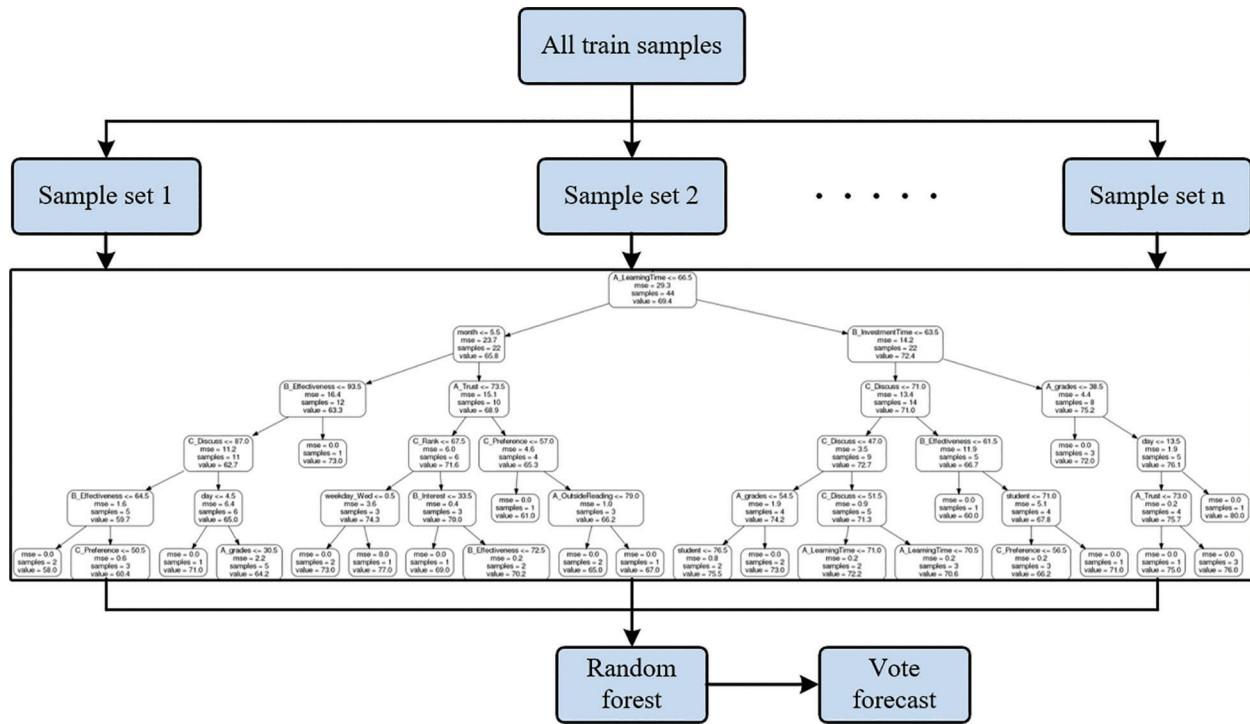


Figure 4: Random forest structure

Table 1: The description of some research variables

Research contents	Variable type	Variable name	Code	Variable assignment
Predicting behavior	Dependent variable	Learning behavior	Result	Bad = 0, good = 1
Subject information	Student information	Admission score	Admission grades	5 points per small drop
		Average study time	Learning time per day	Less than 1 hours per shift
		Extra-curricular reading	Extra-curricular reading quantity	Less than 2 documents per document
	Course information	Optimism state	Optimism state	*
		Professional trust	Professional trust degree	*
		Student's interest	Student's interest	*
		Student input time	Student's devoting time	Every 3 days less downshift
	Teacher information	Importance identification	Importance recognition	*
		Knowledge validity	Knowledge validity	*
		Teacher ranks (consider senior years)	Rank of teacher	Professor = 5, associate professor (5 years and above) = 4, associate professor (less than 5 years) = 3, lecturer = 2, assistant professor = 1
Student's love degree		Student's love degree	*	
Discuss situation with students		Discuss situation with students	*	
Teacher's mental state	Teacher mental state	*		
Student attention	Student attention	*		

(Continued)

Table 1 (continued).

Research contents	Variable type	Variable name	Code	Variable assignment
Course management	Student classroom performance	Teacher-student interaction	Teacher-student interaction	*
		Student-student interaction	Student-student interaction	*
	Student work performance	Assignment submission	Assignment submission	Submit drop files per day for 1 day
		Homework writing	Assignment writing	10 points per small drop
		Homework scores	Assignment score	*
	Student performance	Average scores	Average scores	8 points per small drop
In the two-semester floating		Results change situation in two semesters	To enhance more than 20 points = 5, to enhance more than 10 points = 4, up and down 10 points less than = 3, down 10 points or more = 2, down 20 points or more = 1	
Teaching information	Teacher experience	Textbook	Textbook	*
		Practice	Practice	*
	Teacher observation	Questions	Questions	*
		Comments	Comments	*
	Teacher intuition	Appearance theory	Appearance theory	*
		Internal reflection	Internal introspection	*

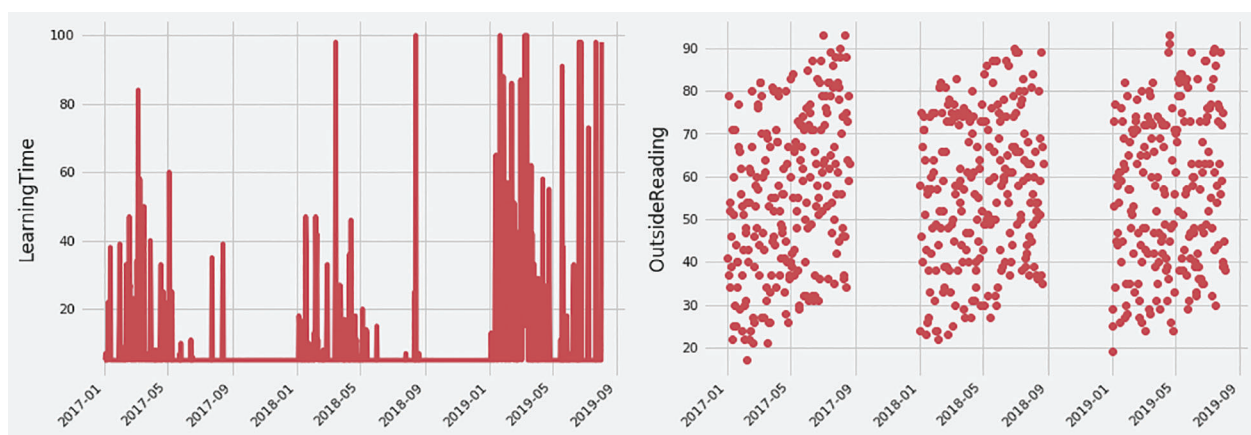


Figure 5: Characteristic data trend

According to the above analysis, the best effect is observed when $m = 6$ and the number of trees is 90. The following 2 experiments are thus carried out with these parameters.

(1) Nine key attributes are extracted from the 19 attributes to obtain the Gini index. The Gini index can be used to judge the importance of the EV (explanatory variable) in a prediction model; meanwhile, the Gini index of a set represents the uncertainty. The larger the Gini index value is, the greater the uncertainty of the samples in a certain class is. This is like entropy. Generally, more information is expected to reduce uncertainty. Therefore, the best way to select a feature partition is to divide the set with the minimum Gini index.

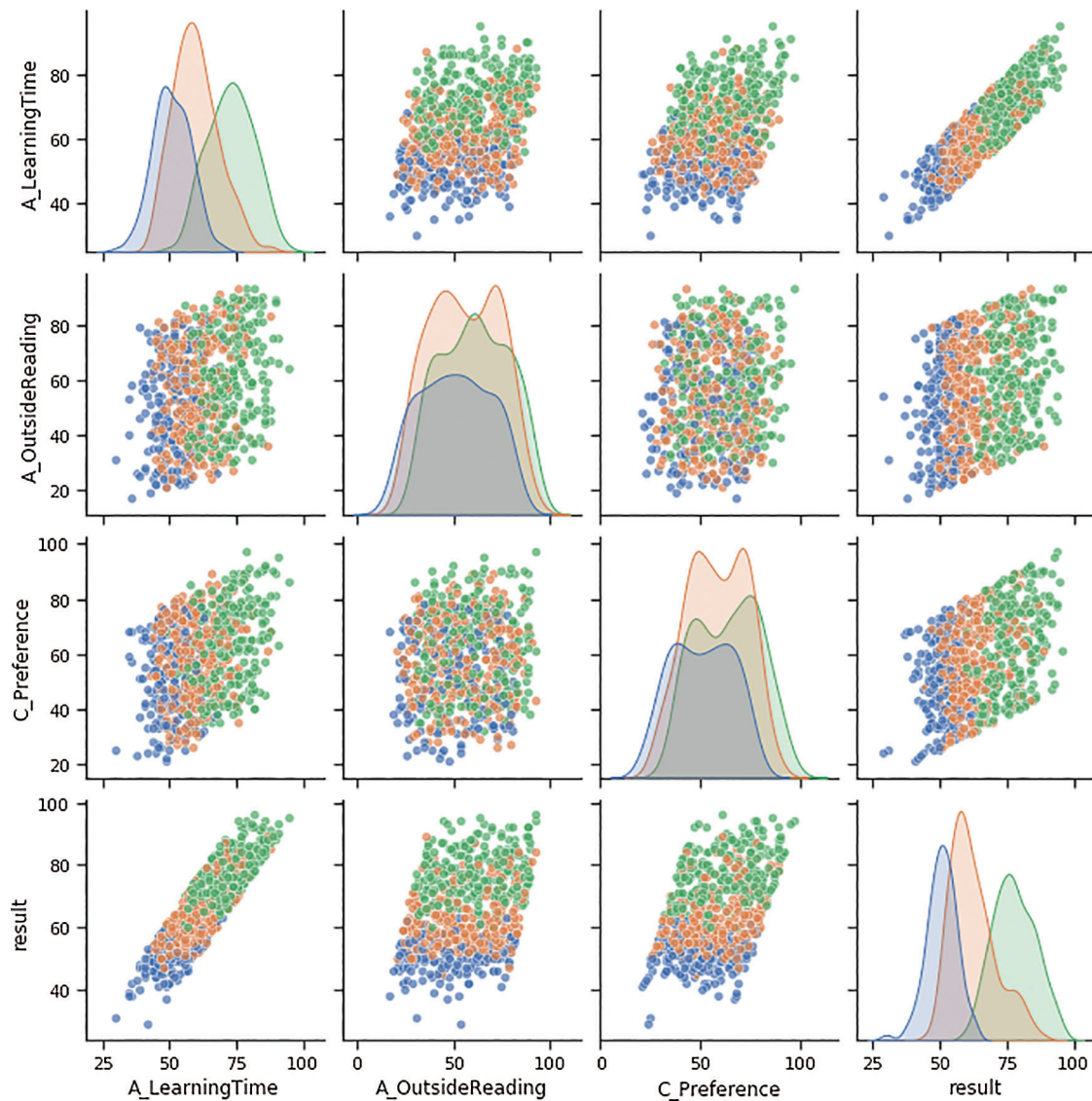


Figure 6: Relationships of four groups of characteristic data

The following [Tab. 2](#) presents the explanation and classification of variables and their importance degrees. The most important variable is the Investment Time, and the weakest is Effectiveness. According to the m obtained from the above experiment, this paper selects the first six variables to predict the learning behavior in the college English teaching model.

(2) A survey of college English teaching

In order to verify the accuracy of the prediction model, which also supports the SVM (support vector machine), CART, and NN (neural network), the 682 valid questionnaires are divided into two groups: 511 to constitute the training sets, and 171 to comprise the test sets. The actual measurement results are shown in [Tab. 3](#).

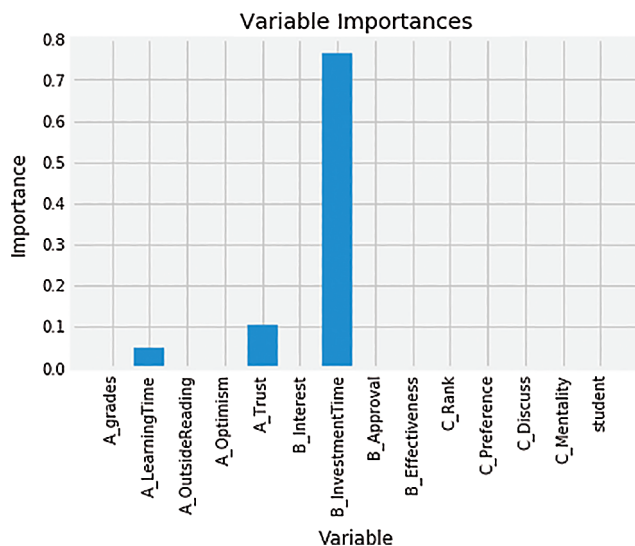


Figure 7: Diagram of feature importance

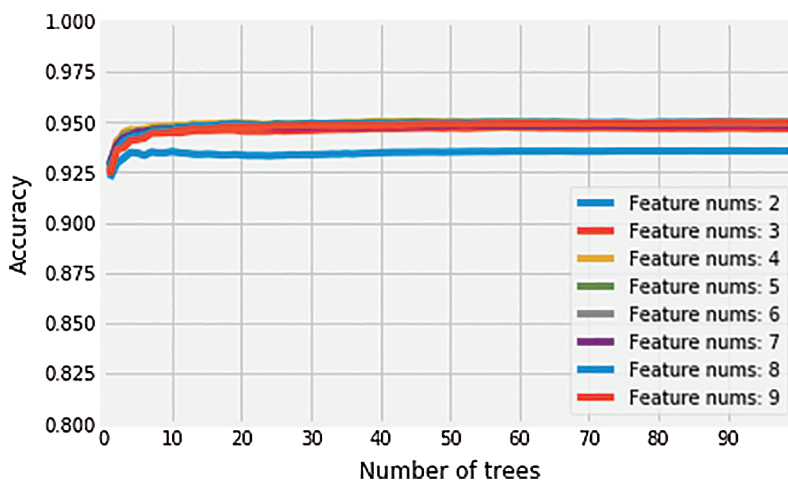


Figure 8: Relationship between different subtrees and accuracy

Table 2: The Gini index values of some variables

EV	Gini	EV	Gini	EV	Gini
Investment Time	0.06	Grades	0.21	Outside Reading	0.26
Trust	0.09	Effectiveness	0.47	Optimism	0.41
Learning Time	0.13	Preference	0.37	Interest	0.17

The above results show that the random forest has good adaptability, discrimination, and generalization ability for learning behavior in college English teaching; moreover, it is obviously better than the other three methods. Therefore, the proposed teaching model based on the attribute prediction is feasible, and the future research on college English teaching model can be guided through the learning behavior prediction method the of random forest.

Table 3: The prediction results of different methods

Method	Training set (%)	Test set (%)	Synthesis (%)
Random Forest	98.4	95.3	97.7
SVM	88.1	79.5	85.9
CART	83.0	80.7	82.4
NN	85.9	84.8	85.6

4 Summary

Using the attributes of teaching entities involved in the college English teaching model, this paper studies an effective analysis and researches on the attribute model of college English teaching and the prediction of students' learning behavior based on the theoretical definition and random forest prediction theory. The feasibility and effectiveness of the scheme are verified by 682 survey data. However, the system still has the following shortcomings:

(1) The analysis of the attributes may not be sufficient. Since only 19 attributes are considered in this scheme, there may be more practical decisions on students' learning behavior, further exploration may enhance the generalization ability.

(2) The time of systematic training and analysis lacks effective analysis. Although the prediction accuracy of this scheme is very high, the training time and decision time cannot be ignored. Classification trees can not only enhance the ability to judge, but also increase the calculation cost; this needs to be further addressed in future works.

(3) The attribute model of college English teaching has been deeply studied, but the follow-up guidance of students' behavior remains to be explored.

To sum up, there are some advancements in the research of the college English teaching model based on attribute prediction analysis, but there remains room for improvement.

Acknowledgement: We would like to thank all the parties involved in this research work.

Funding Statement: This work was financially supported by the Scientific Research Fund of Hunan Provincial Education Department of China (Grant No. 20C1246); and the 2021 Project of Hunan Province Social Science Achievement Evaluation Committee (Grant No. XSP21YBC202).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] D. Stosic, "A functional perspective on the challenges of teaching English tenses to speakers of other languages: The case of adult speakers of Serbian," *Lingua and Education*, vol. 51, no. 1, pp. 79–90, 2019.
- [2] L. Cai, "Japanese teaching quality satisfaction analysis with improved apriori algorithms under cloud computing platform," *Computer Systems Science and Engineering*, vol. 35, no. 3, pp. 183–189, 2020.
- [3] N. Li, "Construction and application of the multi-intermediate multi-media English oral teaching mode," *Intelligent Automation & Soft Computing*, vol. 26, no. 4, pp. 807–815, 2020.
- [4] I. Glogger-Frey, Y. Ampatziadis, A. Ohst and A. Renkl, "Future teachers' knowledge about learning strategies: Misconcepts and knowledge-in-pieces," *Thinking Skills & Creativity*, vol. 28, no. 4, pp. 41–55, 2018.

- [5] A. Zahra, "Current challenges in teaching/learning English for EFL learners: The case of junior high school and high school," *Procedia Social and Behavioral Sciences*, vol. 199, no. 71, pp. 394–401, 2015.
- [6] J. Hong and H. Basturkmen, "Incidental attention to academic language during content teaching in two EMI classes in South Korean high schools," *Journal of English for Academic Purposes*, vol. 48, no. 2, pp. 100921, 2020.
- [7] S. Dimova, "English language requirements for enrolment in EMI programs in higher education: A European case," *Journal of English for Academic Purposes*, vol. 47, no. 3–4, pp. 100896, 2020.
- [8] E. Petek and H. Bedir, "An adaptable teacher education framework for critical thinking in language teaching," *Thinking Skills & Creativity*, vol. 28, no. 3, pp. 56–72, 2018.
- [9] I. H. Erten, "Listening to practising teachers: Implications for teacher training programs," *Procedia Social and Behavioral Sciences*, vol. 199, pp. 581–588, 2015.
- [10] T. Jansen, N. Machts, C. Vgelin, S. Keller and J. Mller, "Judgment accuracy in experienced versus student teachers: Assessing essays in English as a foreign language," *Teaching and Teacher Education*, vol. 97, no. 3, pp. 103216, 2021.
- [11] E. E. Dijk, J. Tartwijk, M. F. Schaaf and M. Kluijtmans, "What makes an expert university teacher? A systematic review and synthesis of frameworks for teacher expertise in higher education," *Educational Research Review*, vol. 31, no. 1, pp. 100365, 2020.
- [12] L. Unsworth and K. A. Mills, "English language teaching of attitude and emotion in digital multimodal composition," *Journal of Second Language Writing*, vol. 47, no. 4, pp. 100712, 2020.
- [13] M. Ryshina, W. Barthold and E. Barthold, "Enhancing the content- and language-integrated multiple literacies framework: Systemic functional linguistics for teaching regional diversity," *System*, vol. 96, no. 2, pp. 102403, 2021.
- [14] D. Rus, "Creative methodologies in teaching English for engineering students," *Procedia Manufacturing*, vol. 46, no. 1, pp. 337–343, 2020.
- [15] S. I. Ciornei and T. A. Dina, "Authentic texts in teaching English," *Procedia-Social and Behavioral Sciences*, vol. 180, pp. 274–279, 2015.
- [16] Z. N. Imam, Z. Nazneen, K. Abdul and A. Shaik, "Developing empathy among first-year university undergraduates through English language course: A phenomenological study," *Heliyon*, vol. 6, no. 6, pp. E04021, 2020.
- [17] T. Kinya, A. Masahiro, A. Maki and O. Yoshiyuki, "Development of educational programs for system creators and business producers in future strategy design in action project group activities through industry-university cooperation," *Procedia Manufacturing*, vol. 39, no. 3, pp. 1377–1382, 2019.
- [18] C. Mihaela, "The challenge of teaching English to adult learners in today's world," *Procedia-Social and Behavioral Sciences*, vol. 197, pp. 1209–1214, 2015.
- [19] J. P. C. Vallente, "Framing pre-service English language teachers' identity formation within the theory of alignment as mode of belonging in community of practice," *Teaching and Teacher Education*, vol. 96, no. 2, pp. 103177, 2020.
- [20] L. Y. Xiang, S. H. Yang, Y. H. Liu, Q. Li and C. Z. Zhu, "Novel linguistic steganography based on character-level text generation," *Mathematics*, vol. 8, no. 9, pp. 1558, 2020.
- [21] Z. Yang, S. Zhang, Y. Hu, Z. Hu and Y. Huang, "VAE-Stega: Linguistic steganography based on variational auto-encoder," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 880–895, 2021.
- [22] S. R. Zhou and B. Tan, "Electrocardiogram soft computing using hybrid deep learning CNN-ELM," *Applied Soft Computing*, vol. 86, no. 4, pp. 105778, 2020.
- [23] C. Arturo, G. S. Claudio and R. C. José, "An online multi-source summarization algorithm for text readability in topic-based search," *Computer Speech & Language*, vol. 66, pp. 101143, 2020.
- [24] L. Xiang, G. Guo, Q. Li, C. Zhu, J. Chen *et al.*, "Spam detection in reviews using LSTM-based multi-entity temporal features," *Intelligent Automation & Soft Computing*, vol. 26, no. 6, pp. 1375–1390, 2020.
- [25] S. Zhou, L. Chen and V. Sugumaran, "Hidden two-stream collaborative learning network for action recognition," *Computers Materials & Continua*, vol. 63, no. 3, pp. 1545–1561, 2020.

- [26] C. M. Yelkanat, "Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm," *Chaos Solitons & Fractals*, vol. 140, pp. 110210, 2020.
- [27] A. K. Tiwary, S. Ghosh, R. Singh, D. P. Mukherjee and P. S. Dash, "Automated coal petrography using random forest," *International Journal of Coal Geology*, vol. 232, pp. 103629, 2020.
- [28] Y. Y. Li, M. Cargill, X. Gao, X. Q. Wang and P. O'Connor, "A scientist in interdisciplinary team-teaching in an English for research publication purposes classroom: Beyond a 'cameo role'," *Journal of English for Academic Purposes*, vol. 40, no. 4, pp. 129–140, 2019.
- [29] H. Zhang, S. J. Jin and S. Z. Du, "Developing a curriculum model of English teaching for master's degree nursing education in a Chinese medicine university," *International Journal of Nursing Sciences*, vol. 7, no. 1, pp. 99–104, 2020.
- [30] C. S. C. Chan, "Investigating a research-informed teaching idea: The use of transcripts of authentic workplace talk in the teaching of spoken business English," *English for Specific Purposes*, vol. 46, no. 2, pp. 72–89, 2017.
- [31] J. Liu, S. Q. Sun, Z. L. Tan and Y. Liu, "Nondestructive detection of sunset yellow in cream based on near-infrared spectroscopy and interval random forest," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 242, no. 1, pp. 118718, 2020.
- [32] Y. Y. C. Ho, "Communicative language teaching and English as a foreign language undergraduates' communicative competence in tourism English," *Leisure, Sport & Tourism Education*, vol. 27, no. 4, pp. 100271, 2020.
- [33] X. M. Li, "Characteristics and rules of college English education based on cognitive process simulation," *Cognitive Systems Research*, vol. 57, no. 1, pp. 11–19, 2019.
- [34] S. R. Zhou, J. Wu, F. Zhang and P. Sehdev, "Depth occlusion perception feature analysis for person re-identification," *Pattern Recognition Letters*, vol. 138, no. 3, pp. 617–623, 2020.
- [35] C. L. Li, X. M. Sun and J. H. Cai, "Intelligent mobile drone system based on real-time object detection," *Journal on Artificial Intelligence*, vol. 1, no. 1, pp. 1–8, 2019.