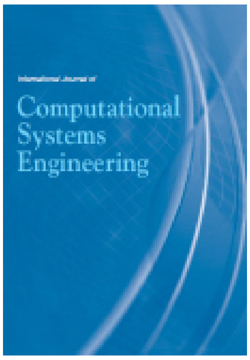


International Journal of Computational Systems Engineering

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

A study on the optimisation of university English teaching based on an enhanced decision tree model in the context of big data

Haiyan Cai

DOI: 10.1504/IJCSYSE.2025.10061164

Article History:

Received:	11 April 2023
Last revised:	26 April 2023
Accepted:	11 June 2023
Published online:	01 August 2025

A study on the optimisation of university English teaching based on an enhanced decision tree model in the context of big data

Haiyan Cai

School of Tourism, Hospitality and Culinary Arts,
Chongqing City Vocational College,
Chongqing Yong Chuan, 402160, China
Email: haiyan_cai@outlook.com

Abstract: The purpose of learning is the most important factor affecting an individual's percentage of making notable progress (PMNP). In the context of big data, this study selected a sample of 1,805 non-English majors in a university to investigate the optimisation of university English teaching based on an enhanced decision tree model. The results showed that among the 1,805 student sample, 352 students made more significant improvements in their test scores compared to the last test, accounting for 19.50% of the total PMNP of the entire sample. The Chi-squared automatic interaction detector (CHAID) decision tree model was used to identify implicit and valuable factors influencing teaching quality based on data on the process, conditions and environment of English language teaching. The results show that through calculation of CHAID decision tree, the resultant data of each node is a reflection of the effect of each factor on PMNP.

Keywords: big data; CHAID decision tree; English language teaching; percentage of making notable progress; PMNP; genetic algorithm.

Reference to this paper should be made as follows: Cai, H. (2025) 'A study on the optimisation of university English teaching based on an enhanced decision tree model in the context of big data', *Int. J. Computational Systems Engineering*, Vol. 9, No. 12, pp.1–11.

Biographical notes: Haiyan Cai obtained her BE in English from Chongqing University of Posts and Telecommunications in 2006. She obtained her ME in English Teaching from Chongqing Normal University in 2012. Presently, she is working as an Associate Professor in the School of Tourism, Hospitality and Culinary Arts, Chongqing City Vocational College. Her areas of interest are English education and vocational education.

This paper was originally accepted for a special issue on 'Educational Innovation and Sustainability in the Digital and Intelligent Era' guest edited by Assoc. Prof. Giovanni Pau.

1 Introduction

With the increasingly rapid development of the internet, the economy and life of human society have entered a new era of digitisation, informatisation and globalisation (Xu et al., 2017; Hsieh et al., 2016; Hung, 2015), which makes English increasingly important. As a communication tool, English is one of the essential language skills for elite people in society today (Alsowat, 2016; Basal, 2015; Wickramarachchi et al., 2016). As one of the world's leading internationally spoken languages today, English has been widely used in all areas of human life (Spieler and Miltenberger, 2017). Many countries around the world now have English education as an important subject. Most schools have incorporated English education into the basic education curriculum. English education has become an important part of the quality education of citizens and is in a very important position (Kvashnina and Martynko, 2016). English is one of the compulsory subjects during secondary school, university and postgraduate studies in China. The English IV and English VI exams have an important part in the various university exams. Thus, it is very important for university students to learn English well (Mashael and Muna, 2016). In recent years, with the emergence of Internet+, all walks of life have undergone great changes. Internet+ education has gradually become a development trend, and people are paying more and more attention to online education platforms (Menegazzo, 2017). Through online English learning platforms, more students are willing to take the initiative to carry out independent learning (Hung, 2017). In English teaching, university teachers can train students in various aspects such as speaking, listening, writing and reading through online English platforms, which greatly improves students' learning efficiency and motivation (Eva, 2015; Lee, 2017; Schneider et al., 2017).

Data mining techniques are widely used in a variety of industries and have yielded certain results. We use an online English learning platform to mine and analyse student learning data. The results of model training were used to explore the correlation between English language test scores and various factors (Handoyo, 2016), which is of great value to both teachers' teaching and students' learning.

Therefore, it is essential to apply data mining techniques to the teaching of English at university (Baum and Boughton, 2016). According to the application areas, educational data mining can be classified into teaching management applications, teaching research applications, research management applications and so on (Chuang et al., 2018; Mari and Mariko, 2015; Mancuso and Miltenberger, 2016). Data mining can mine potential and useful knowledge from a large amount of data.

Machine learning is currently one of the core research areas of artificial intelligence and one of the main approaches to solving data mining problems. By training large datasets, machine learning allows data mining tasks to be accomplished using a variety of relevant algorithms (Chen et al., 2016; Chodkiewicz and Miniszewska, 2015; Doman and Webb, 2016; Ramos, 2015; Şahin, 2020; Ward 2016). The decision tree algorithm is the dominant classification algorithm in machine learning (Candel-Mora, 2015). By training a dataset with labels, the decision tree algorithm can produce a binomial or multinomial tree. For a decision tree, each internal node represents a test on an attribute, while each branch is a test output. The nodes of each tree leaf represent classes or class distributions (Gerbensky, 2017; King and Finn, 2017; Lertworaprachaya et al., 2014). Currently, the construction of decision trees needs to be performed using pruning for detection and to cut out noisy, isolated points from the training data, so that the accuracy of decision tree classification can be improved on unknown data. In order to solve this problem, this paper proposes an enhanced CHAID decision tree based on genetic algorithm. Based on this, this paper conducts a study on the optimisation of university English teaching based on the enhanced decision tree model in the context of big data.

2 Related work

At this stage, the research on teaching optimisation in the context of big data mainly includes three aspects: literature search, theoretical research and experimental methods. First, in the field of university English teaching, to understand the current development of data mining, Lee and Wallace (2018) conducted a search for appropriate keywords in mainstream literature databases and retrieved computer-assisted languages. After screening the key literature, aspects of the research were categorised, leading to a literature review. The literature was sourced from journal networks such as Wanfang and China Knowledge Network (CNKI), and also included information from search engines such as Baidu and Google. Abbasi and Izadpanah (2018) used the literature analysis method to provide an effective understanding of the current state of development of data mining in countries where English is a second foreign language. After taking a holistic view of the development of the field, it is pointed out that the purpose of learning is the most important factor influencing individual percentage of making notable progress (PMNP). However, the factors that influence individual PMNP cannot only include the purpose of learning, and more factors should be analysed.

Secondly, many researchers have conducted theoretical studies on the application of data mining techniques in teaching. Freiermuth (2017) analysed the general process of data mining, the main tasks, and described the common feature selection methods and common optimisation algorithms for machine learning, such as ID3 decision trees and C4.5 decision trees. Finally, the experimental method belongs to a very important research approach and occupies a very important place in scientific research. Komiya et al. (2017) verified the feasibility of ID3 decision trees in teaching optimisation by means of an experimental method. Experiments are thus a bridge between theory and practice (Le et al., 2015).

In this paper, a user profile analysis is first conducted to help teachers gain a better understanding of their students' basic profiles. Then, a decision tree model is selected to conduct a rational experimental design and to analyse the main factors affecting the PMNP of the student population.

The main innovations and contributions of this paper include.

1 In addition to the learning purpose, the main factors that affect the PMNP of the student group are expanded and analysed, including learning motivation, teaching means and teaching mode, etc.

2 In order to improve the accuracy of decision tree classification on unknown data, an enhanced CHAID decision tree based on genetic algorithm is proposed. The enhanced CHAID decision tree model is used to mine the factors influencing teaching quality and rank their importance.

3 Big data analysis based on decision tree principle

3.1 Overview of the decision tree algorithm

The decision tree classification algorithm is a type of algorithm that performs unordered training data classification. After the decision tree has been constructed, the final classification result can be obtained by multiple judgments under the given conditions. First, the data is input at the root node of the decision tree. Then, at each node, the branching conditions are judged. Subsequent branches are selected according to the judgement until a leaf node is reached. Each leaf node is a representative of a category. The data category is the last leaf node selected. When constructing a decision tree, it is necessary to first determine the cut-off conditions for each attribute of the nodes and then to set up the tree structure between them. The information entropy can determine the conditions for the construction of the decision tree (Mercedes et al., 2019).

$$info(D) = -\sum_{i=1}^{m} pi \log_2 \left(p_i \right) \tag{1}$$

where p_i denotes the probability of the i^{th} category occurring in the entire training data. In order to maximise the classification effect of the decision tree, the parameter with the highest information entropy needs to be selected during the selection of the classification parameters (Suranakkharin, 2017). In the construction of decision tree, over-fitting problem can be avoided by pre-pruning.

3.2 CART decision tree algorithm

The first decision tree algorithm is the Iterative Dichotomizer3 (ID3) algorithm, and most of the subsequent proposed decision tree algorithms are based on the refinement and improvement of the ID3 algorithm. ID3 algorithm uses divide-and-conquer strategy to classify data, and uses information entropy gain value to build judgment conditions.

$$\Delta info(D) = info(D_1) - info(D_2) \tag{2}$$

Based on the attributes of each node, the maximum category information of all records is obtained when testing each non-leaf node. The advantages of the ID3 algorithm are high learning ability, simple method and fast classification. The ID3 algorithm is suitable for large-scale simple data processing problems. The disadvantage of the ID3 algorithm is poor noise immunity. When dealing with data with multiple attributes and large entropy values, it is difficult for the ID3 algorithm to find a suitable judgement result.

After improving the ID3 algorithm, the C4.5 algorithm was obtained. Compared to the ID3 algorithm, the C4.5 algorithm has an information gain rate.

$$\Delta info(D) = \Delta info(D_1) / \Delta info(D_2) \tag{3}$$

When selecting data that can be used in place of information gain values, it is necessary to keep the attribute partitioning unbiased towards multi-valued attributes. To avoid overfitting, prepruning is required when constructing the C4.5 decision tree, which has reduced time complexity and is therefore effective in reducing program runtime.

Unlike the ID3 and C4.5 decision trees, the CART decision tree algorithm can be used for both classification and regression problems. Since the CART decision tree uses binary recursion to divide the sample set into two subsamples, it generates a total of two branches of non-leaf nodes. The CART decision tree generates a binary decision tree with two options (yes or no) for each level of decision. The CART decision tree algorithm has the advantage of being able to work with continuous fields. The CHAID has functions such as variable filtering, target selection and cluster analysis. It is suitable for tasks such as classification and ranking.

Compared with other decision tree algorithms, CHAID algorithm has the following advantages in big data mining:

1. CHAID algorithm can deal with multi-classification problems, while other decision tree algorithms such as CART algorithm are only suitable for binary classification problems or regression problems.
2. CHAID algorithm uses chi-square test to select features, which can consider the cross relationship between variables, while other decision tree algorithms such as ID3 and C4.5 only consider the information gain or information entropy of a single feature.

Therefore, CART decision trees are chosen to implement data mining tasks in the context of big data.

4 Experimental design and sample characteristics selection

4.1 Big data acquisition

Using SSH framework technology and Docker container technology to build a big data platform, and using Python to capture students' real data. In this paper, 1,805 non-English majors in a university were selected for a survey between March 2 and May 2, 2022, and the CHAID decision tree algorithm was used to analyse the factors influencing teaching quality. The objective evaluation index was the PMNP of individual students. The sample was optimally split according to the screened explanatory and response variables. According to the significance of chi-square test, the multivariate list is automatically grouped and judged.

4.2 Modelling of various types of students

The data set to be classified is assumed to be $X = \{X_1, X_2, \cdots, X_m\}$. A set of data elements is denoted by $X_i (i = 1, 2, \cdots, m)$ and *m* denotes the size of a set of data elements. The set of attributes of a data element is denoted as $P = \{P_1, P_2, \cdots, P_m\}$. $P_i (i = 1, 2, \cdots, k)$ is some attribute and the length of the attribute is *k*. The students' values for the five cognitive ability tests are shown in Table 1. '1' indicates correct, '–l' indicates incorrect, and '0' indicates abstention.

Table 1 Students' values for the five cognitive ability tests

	Comprehension capacity	*Memory capacity*	*Analytical capacity*	*Application capacity*	*Evaluation capacity*
1	–1	–1	0	1	–1
2	0	0	1	–1	–1
3	0	1	0	–1	0
...	...	...	...	...	...
N	1	1	–1	0	1

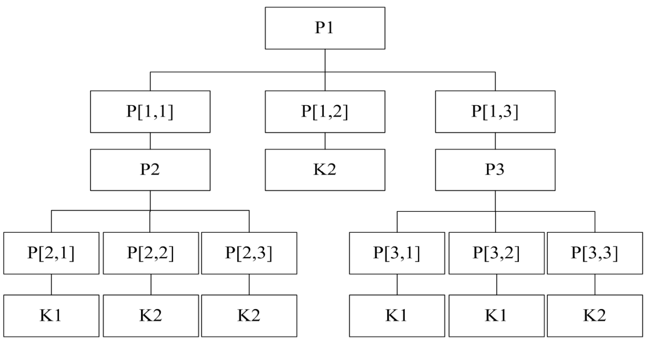
Table 2 A sample of the limited training set

P1	P[1, 3]	P[1, 2]	P[1, 1]	P[1, 2]	P[1, 1]	P[1, 1]	P[1, 1]	P[1, 1]	P[1, 1]	P[1, 1]
P2	P[2, 3]	P[2, 1]	P[2, 1]	P[2, 2]	P[2, 2]	P[2, 1]	P[2, 1]	P[2, 1]	P[2, 1]	P[2, 1]
P3	P[3, 3]	P[3, 2]	P[3, 1]	P[3, 2]	P[3, 2]	P[3, 2]	P[3, 1]	P[3, 1]	P[3, 1]	P[3, 1]
P4	P[4, 2]	P[4, 3]	P[4, 2]	P[4, 3]	P[4, 3]	P[4, 4]	P[4, 2]	P[4, 1]	P[4, 1]	P[4, 1]
P5	P[5, 2]	P[5, 1]	P[5, 1]	P[5, 2]	P[5, 3]	P[5, 3]	P[5, 3]	P[5, 2]	P[5, 1]	P[5, 1]
Type	*K2*	*K2*	*K1*	*K2*	*K1*	*K2*	*K2*	*K2*	*K1*	*K1*

The classification decision tree is a predictive model that requires the use of a finite set $S(\subseteq D)$ to perform the necessary model training. Table 2 shows a sample finite training set.

Figure 1 shows the decision tree corresponding to the finite training set. It can be seen that any node at the top of the decision tree represents the test result for an attribute. Each node continues down the selection process, indicating some likelihood of an attribute occurring.

Figure 1 Decision tree



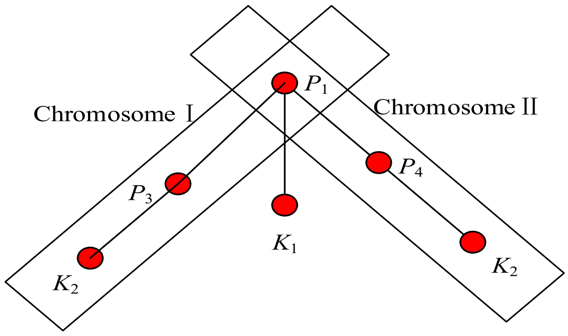
In a decision tree, each join of the root node represents an if-then classification pattern. X is usually a massive database in practical application, with the order of magnitude up to GB and TB. Thus, $X = E1 \cup E2 \cup \cdots \cup En$, n is taken to infinity.

4.3 *Enhanced CHAID decision tree based on genetic algorithm*

In the construction of decision trees, the use of pruning is required to detect and cut out the noisy, isolated points from the training data, so that the accuracy of decision tree classification can be improved on unknown data. To solve this problem, an enhanced CHAID decision tree based on genetic algorithm is proposed in this paper. Each branch in the propose enhanced decision tree is represented by a chromosome in the genetic algorithm (Amerian and Tajabadi, 2020). Genetic algorithm can be used to ensure that tall fitness survive. The weight adjusts the contribution of each feature to the fitness function. Finally, the proposed enhanced decision tree employs proportional selection to compute undetermined individuals.

The attribute categories are denoted by P_i and the categories predicted by the data are denoted by K_i. The chromosome individuals are shown in Figure 2.

Figure 2 Stained individual (see online version for colours)



The chromosomal gene sequence is $P_1, P_2, \cdots, P_k, K_i$. Gene values indicate a certain attribute. Gene values are expressed as binary values of three digits, as shown in Table 3. Gene values are the results of tests for an attribute.

Table 3 Coding of gene values

	P1	*P2*	*P3*	*P4*	*P5*	*P6*
Attribute test values (genetic values)	111	110	101	010	111	001

To perform the search out all available solutions, a completely randomised approach is used to represent the initial population in the genetic algorithm (Hellwig, 2022). For decision trees, the set of each tree representation is based on the if-then model, which also holds for chromosomes. Thus, randomly constituting the set of decision trees and the initial population have the same meaning. The fitness function is a polynomial summation of weights.

$$F(X) = w_4(1 - g_4) + g_3 w_3 + g_2 w_2 + g_1 w_1 \tag{4}$$

where g_1 denotes the effective gene length of P, g_2 denotes the full effective gene length of X, g_3 denotes the classification rule of X, g_4 denotes the proportion of misclassification of X, and w_i denotes the weights.

In this paper, the random residual selection method is used. First, the survival period of the offspring of each individual in the population is calculated.

$$N_i = M \times \left(F_i \Big/ \sum_{i=1}^{M} F_i \right), (i = 1, 2, \cdots, M) \tag{5}$$

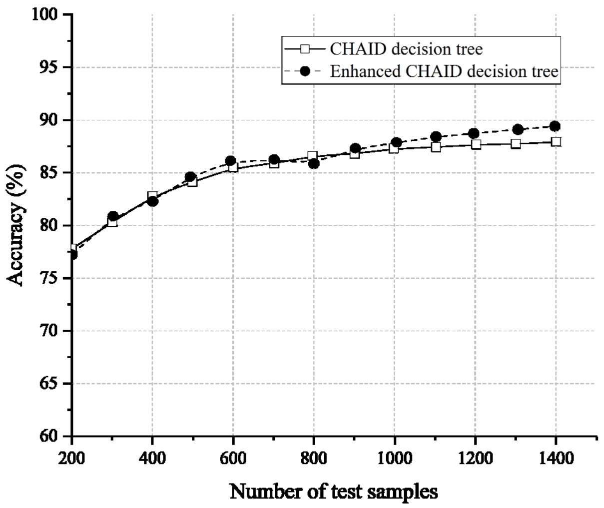
Denote by $[N_i]$ the value of N_i after rounding to obtain the individual fitness.

$$F_i - [N_i] \times \sum_{i=1}^{M} (F_i / M) \tag{6}$$

4.4 Performance verification of enhanced CHAID decision tree

In order to verify the performance of the proposed enhanced CHAID decision tree algorithm, the UCI machine learning data set is used for classification test verification. A total of 2400 samples were tested. The training samples are 1,000 randomly selected samples and the rest are test samples. Each experiment was repeated 10 times and the average value was removed as the final result. The comparison of classification accuracy between the standard CHAID decision tree algorithm and the enhanced CHAID decision tree algorithm is shown in Figure 3.

Figure 3 Comparison of classification accuracy between the two algorithms



As can be seen from Figure 3, with the increasing number of test samples, the classification accuracy of the enhanced CHAID decision tree algorithm is continuously improved, which is because the genetic algorithm can ensure the survival of people with higher fitness.

4.5 Sample feature selection

The sample data was obtained from the learning data generated by the University English Assessment System. In addition, a questionnaire was used to conduct the research with 857 female students and 948 male students. A total of 1850 questionnaires were distributed and 1805 valid questionnaires were received. The return rate of valid questionnaires was 97.56%. In order to effectively analyse the factors influencing the quality of teaching and learning, a variety of possible attributes such as the means of teaching, the purpose of learning, the type of teacher and the mode of teaching were designed into the questionnaire. These attributes may reflect subjective factors in student learning, such as teacher evaluation and purpose of learning. Others reflect objective factors such as teaching conditions and environment, including teaching means, teaching mode, teacher qualifications, etc. The sample characteristics were selected as shown in Table 4.

Table 4 Selection of sample characteristics

Property name	*Property values*	*Number of samples*	*Proportion of total sample (%)*
Gender	Male	948	52.52
	Female	857	47.48
Teaching tools	Multimedia teaching	1145	63.43
	Non-multimedia teaching	660	36.57
Learning objectives	Passing the exam	628	34.79
	Interest	570	31.58
	Find a job or go abroad	607	33.63
Teaching mode	Happy English	666	36.89
	University English (A)	442	24.49
	University English (B)	231	12.79
	New Horizons English	191	10.58
	English for High Starters	185	10.25
	Other	90	5.00
Overall student evaluation of the teacher	Excellent	1,039	57.56
	Good	693	38.39
	Medium or poor	73	4.05
Student-teacher communication outside the classroom	More	328	18.17
	General	337	18.67
	Less	1140	63.16
Qualifications of the teachers who teach the courses	PhD	207	11.47
	Masters	1,465	81.16
	Undergraduate	133	7.37
Students' evaluation of the teaching materials	Good	743	41.16
	General	852	47.20
	difference	210	11.64
Language of instruction	English-only teaching	675	37.39
	Bilingualism	1,130	62.61
Research level of teachers	High	448	24.82
	General	1,128	62.49
	difference	229	12.69
Students' comments on the difficulty of the examination	It is difficult to be on the other side of the fence.	462	25.59
	Moderate	1,068	59.17
	easy to do	275	15.24

Table 4 Selection of sample characteristics (continued)

Property name	*Property values*	*Number of samples*	*Proportion of total sample (%)*
Entry English grades	Top 30%	550	30.47
	Top 30%-60%	592	32.79
	Other	663	36.74
Type of teacher	Foreign teachers	282	15.62
	National teachers	1,523	84.38
Total		*1,805*	*100*

4.6 User analysis

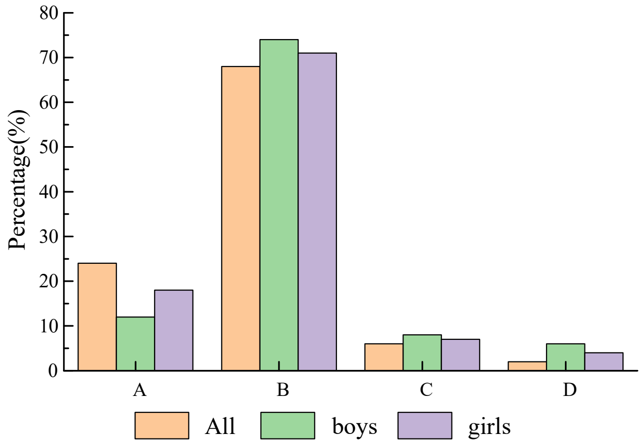
4.6.1 The main purpose of studying English during university

The results of the data analysis show that improving general quality is the main purpose of learning English. The reason for this is that most students believe that learning English well will help them to be competitive when looking for a job or going abroad. As far as interest in learning English is concerned, girls prefer learning English to boys. Some of the students' study for the purpose of gaining a certain number of credits to complete their studies. In the case of unclear purposes of study, the proportion was significantly higher for boys than for girls. In addition, 1.39% of the students studied for the purpose of receiving praise from their teachers. It is clear from the data that students are more likely to agree on the importance of English as a skill in their personal future development, but are less interested in learning English. The main purposes of learning English during university are shown in Table 5.

4.6.2 Attitudes towards learning English

Students can be classified into four types based on their attitude to learning. Students of Type A will occasionally actively seek opportunities to learn English. Students of Type B regularly seek out opportunities to practise English on their own initiative. Students of Type C never actively learn English in class. The analysis of the data showed that 69.25% of students actively seek opportunities to learn English occasionally, with a lower proportion of girls than boys. 1.33% of students chose other. Some of these students felt that there was no ideal environment for them to learn English, and some were unwilling to learn English. A higher proportion of boys than girls chose other. It can be seen that most students occasionally actively seek opportunities, which means that most students do not have the habit of actively learning English. Attitudes towards learning English are shown in Figure 4.

Figure 4 Attitudes towards learning English (see online version for colours)



4.6.3 The most important skills in English

The main skills in the process of teaching English are divided into reading, listening, writing, translation and others. As can be seen from Table 6, 77.95% of boys consider listening to be important, compared to 92.41% of girls. In reading, the proportion of boys was significantly higher than that of girls. Very few students thought that translation was important. In writing, the proportion of boys who thought writing was important was 1.37% and the proportion of girls who thought it was important was 1.40%. It can be seen that most students consider English listening and speaking skills to be more important than skills such as translation and reading. This shows that many students have dumb English in the process of learning English, so strengthening training in listening and speaking is the key to learning English well.

Table 5 Main purposes of studying English during university

	Liked (%)	*Communication needs (%)*	*Improving overall quality (%)*	*Working or going abroad (%)*	*No purpose (%)*	*Completion of studies (%)*	*Family expectations (%)*	*Praise from teachers (%)*	*Other (%)*
All	12.52	10.58	33.57	26.70	4.87	4.98	1.22	1.39	4.17
Boys	11.28	10.97	34.59	24.58	5.91	8.23	1.58	1.37	1.49
Girls	13.54	11.90	31.04	31.97	2.80	3.03	1.45	1.98	2.29

Table 6 Most important skills in English

Skills	*Heard (%)*	*Reading (%)*	*Translation (%)*	*Writing (%)*	*Other (%)*
All	85.15	11.34	1.22	0.55	1.74
Boys	77.95	16.24	3.16	1.37	1.28
Girls	92.41	2.45	1.05	1.40	2.69

4.6.4 Time spent learning English outside the classroom

The amount of time spent studying English outside the classroom was divided into about 1 hour, about 2 hours, more than 3 hours, almost zero and other. As can be seen from Table 7, the majority of students, 43.76%, spent around 1 hour studying English. Very few students studied English for about 2 hours and even fewer for more than 3 hours. 2.45% of the girls chose other, and this group of students studied English for an irregular amount of time.

Table 7 Time spent learning English outside the classroom

Time	*Greater than 3 hours (%)*	*Around 2 hours (%)*	*Around 1 hour (%)*	*Almost zero (%)*	*Other (%)*
All	5.21	10.19	43.76	38.34	2.50
Boys	3.69	11.39	45.15	38.29	1.48
Girls	8.17	12.48	46.21	30.69	2.45

4.6.5 Ways of learning English after school

There are five ways of learning English, such as reading English novels etc., watching English movies and writing English emails. As can be seen from Table 8, 56.12% of the students learn English by watching English movies or English TV series. The proportion of boys learning English by reading magazines or speaking is higher than that of girls. Few students chose English corner or English mail to learn English. Students who chose other ways to learn English mainly through English dictionaries and test papers.

4.6.6 Students' expectations of learning English in context

Students' expectations of learning English in context were investigated according to three different teacher compositions, as shown in Table 9.

The statistics and categorisation of the data show that over 80% of students want to improve their English language skills. Over 90% of students want to improve their English listening and speaking skills. More than 70% of students would like to see more interaction between teachers and students, thus creating a relaxed classroom atmosphere. Students would like to see more opportunities for foreign teachers to interact with each other in class. For the co-teaching model with Chinese and foreign teachers, students would like to enrich the content and increase their interest in learning English. For the Chinese teacher model, students would like the teacher to be more lively and to enhance the teaching of key content. For the foreign teacher model, students would like to see a greater variety of teaching methods, such as more time spent on Western culture (Wang et al., 2017). Students would like to see the number of foreign teachers increase and for foreign teachers to focus on students with poor listening and speaking skills (Perry, 2015).

Table 8 Ways of learning English outside the classroom

Mode	*Reading fiction (%)*	*Go to English corner (%)*	*Writing e-mails (%)*	*Watching movies (%)*	*Othe r (%)*
All	12.46	13.71	9.92	56.12	7.79
Boys	14.56	6.12	6.75	69.73	2.84
Girls	11.43	7.12	6.42	71.88	3.15

Table 9 Students' expectations of contextual English learning

Cooperation between Chinese and foreign teachers	*Teachers in China*	*Foreign teachers*
Hope it works well	Would like a more lively classroom atmosphere	I hope to exchange more
Complementary strengths	More emphasis on important content	More about foreign cultures
Improve English	Easy communication	Practising listening and speaking skills
Improving listening and speaking skills	Give more time to practice speaking	Talk more with students
Teach us some oral sentence	Introduce some language back ground	nothing
I hope this way could promote my speaking	Encourage class to speak more	Correct my pronunciation

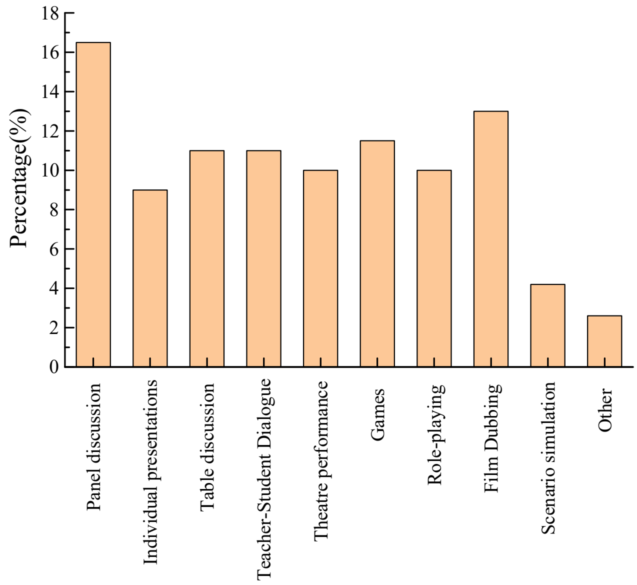
Table 10 Reasons for learning English well

Reasons	*Individual talent (%)*	*Teaching conditions (%)*	*Parental help (%)*	*Teacher level (%)*	*Textbooks (%)*	*Other (%)*
All	53.46	35.78	3.10	1.77	1.99	3.90
Boys	51.47	35.65	3.90	3.48	1.89	5.61
Girls	54.14	30.69	4.08	1.98	3.73	3.38

4.6.7 Reasons for learning English well

The reasons for learning English well were divided into six categories according to the teaching conditions, personal talent, parental help, teacher level, teaching materials and others. As can be seen from the data in Table 10, 53.46% of the students considered better personal talent as the main reason and a key factor in learning English well. 35.78% of students considered the learning environment and teaching conditions to be the main factors. 3.10%, 1.77% and 1.99% of students thought that parents' help, teachers' level and teaching materials were the main reasons. Among the students who chose other reasons, there were more boys than girls. This section of students considered attitude towards learning English, interest and effort as the main factors for learning English well.

Figure 5 Ways to improve speaking skills (see online version for colours)



4.6.8 Ways to improve speaking skills

As can be seen from Figure 5, the largest number of students, 16.51%, thought that the group discussion approach could work on speaking. 13.02% of students thought that film dubbing could effectively improve speaking skills. 11.19% of students thought that role-playing could effectively improve speaking skills. 11.02% of students thought that table discussions were effective in improving speaking skills. 10.91% of students thought that teacher-student dialogues were effective in improving speaking skills. The students who chose the others mainly chose the appropriate practice method according to their own situation.

4.6.9 Ways in which group discussions promote speaking skills

The ways in which group discussions promote speaking skills are shown in Table 11. It can be seen that 55.57% of the students thought that the most effective way was to follow a discussion on a pre-determined topic without any role. 38.17% of the students thought that the most effective way was to take turns in leading the discussion on their respective topics of their own choice. 3.77% of students thought that the most effective way was to take turns to speak in order as a class. Other views were expressed by 2.49% of students.

Table 11 Ways in which group discussions promote speaking skills

Mode	*Pre-set topics*	*Self-selected topics*	*The class will take turns to speak in turn*	*Other*
Number of people	1,003	689	68	45
Percentage/ %	55.57	38.17	3.77	2.49

Figure 7 PMNP decision tree (partial)

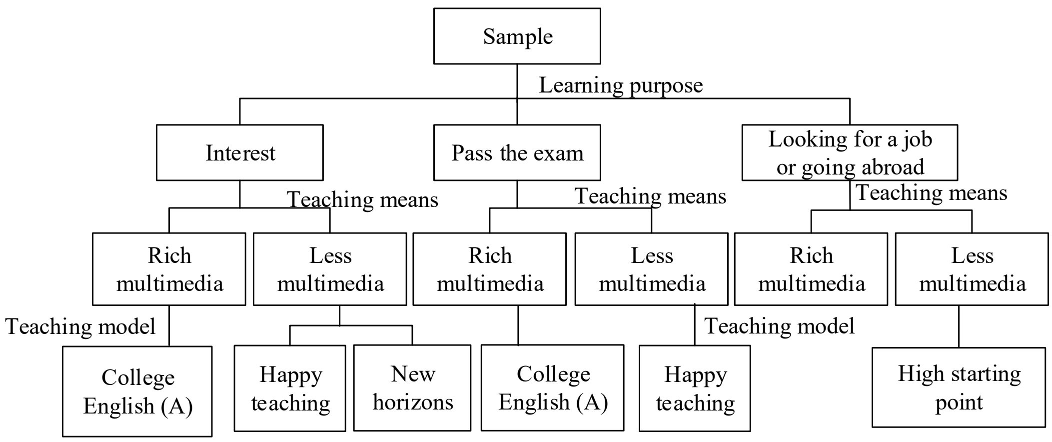


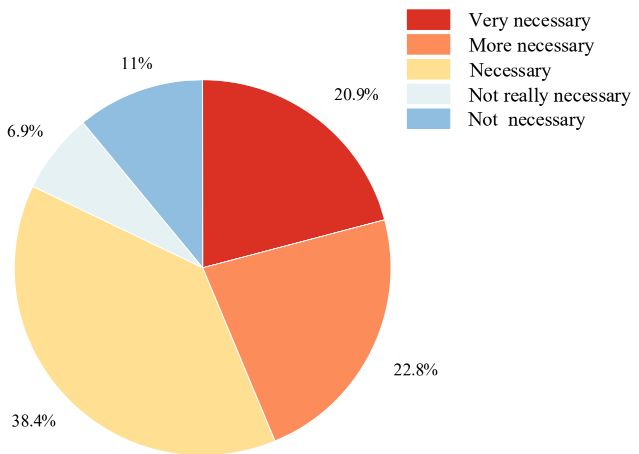
Table 12 Acceptance of practice questions

	Very adaptable (%)	*Comparative adaptation (%)*	*Fairly good (%)*	*Maladjustment (%)*	*Very uncomfortable (%)*
Reading	29.58	40.78	22.83	4.21	2.6
Listening	19.50	36.18	35.18	7.31	1.83
Writing	12.57	31.52	41.22	10.58	4.11
spoken language	7.42	22.49	33.91	29.36	6.82

4.6.10 The need for follow-up reading

For the listening and speaking part of the after-school work, 82.11% of the students thought that it was necessary to follow along. In terms of English learning, English is an audible language and reading along is useful for speaking practice and can improve English listening skills significantly. As can be seen from Figure 6, the majority of students support the practice of reading along.

Figure 6 The need to follow the reading (see online version for colours)



4.6.11 Acceptance of practice questions

In terms of the acceptability of different types of practice questions, 93.19% of the students were comfortable with reading practice questions, 90.86% with listening practice questions, 63.82% with writing practice questions and 85.31% with speaking practice questions. Table 12 shows the level of acceptability of the practice questions.

5 Experimental results and their analysis

5.1 Importance of factors affecting the quality of learning

The analysis results of the enhanced CHAID algorithm are expressed in the form of PMNP decision tree, and the specific steps are as follows:

1 Data preprocessing: preprocessing the original data, including missing value processing, abnormal value processing, data cleaning, data normalisation, etc.

2 Building decision tree: building the enhanced CHAID decision tree according to the selected features. In CHAID algorithm, the tree is constructed according to chi-square test and difference exclusion method, and the decision tree is split according to the standard of optimal segmentation until the termination condition is reached.

3 Decision tree pruning: In order to avoid over-fitting, it is necessary to prune the constructed decision tree to improve the generalisation ability of the model.

4 Model evaluation: use the test set to evaluate the built decision tree model.

5 Model application: Apply the enhanced CHAID decision tree model to big data mining, and make predictions and decisions according to business requirements. At the same time, it is necessary to monitor and optimise the output results to avoid problems such as model over-fitting and abnormal data.

The decision tree has 55 nodes at the 95% confidence level, of which 15 significant nodes have been selected for this study. Figure 7 shows the PMNP decision tree.

Table 13 Statistical indicators for selected nodes

Node number N	*A*	*B(%)*	*C*	*D(%)*	*PMNP (%)*	*E*
8	43	2.38	14	3.98	32.56	1.67
6	17	0.94	4	1.35	23.53	1.44
5	260	14.40	70	19.89	26.92	1.38
1	494	27.37	121	34.38	24.49	1.26
4	285	15.79	59	16.76	20.70	1.06
3	162	8.98	29	8.24	17.90	0.92
21	76	4.21	15	4.26	19.74	1.01
18	12	0.66	3	0.85	25.00	1.29
9	108	5.98	17	4.83	15.74	0.81
14	141	7.81	21	5.96	14.89	0.76
13	206	11.41	22	6.25	10.68	0.54

The quality of learning in this study is represented by PMNP (Stob, 2015). The degree of importance of each influencing factor on PMNP can be found by augmenting the CHAID decision tree (Reinhardt, 2020). In the case of the enhanced CHAID algorithm, the multivariate list is automatically judged to be grouped according to a chi-square test of significance. For the influences derived from the classification results, the closer to the root of the decision tree the more important they are (Ahmadi and Reza, 2018). Of the 1,805 student sample, 352 students made more significant improvements in their exam results compared to the last exam, representing a total PMNP of 19.50% of the entire sample. The purpose of study is the most important factor influencing PMNP. For example, the PMNP of students whose purpose of learning was interest was nearly twice as high compared to those who passed the examination. According to the classification order of the decision tree, teaching methods were the second most important factor influencing PMNP, followed by teaching mode and other factors.

5.2 Analysis of the effect of each factor

In an enhanced CHAID decision tree, each node is given a specific classification group. This outcome data is a reflection of the effect of each factor on PMNP and can be described in an if...then... form. For example, node 14 can be described as reflecting the rule: If (motivation for learning is interest) and (teaching method is multimedia) and (teaching mode is English for pleasure), then PMNP = 14.5%. By analysing the data from each node in the decision tree, it is possible to derive the effect of each factor on PMNP, including two aspects. Table 13 shows the statistical indicators for some of the nodes. Firstly, a visualisation of the teaching methods, teaching modes and teachers' qualifications. Secondly, the statistics of each node can quantitatively show the specific effect of each influencing factor. For example, of the 43 samples included in node 8 in Table 13, 14 samples showed significant improvement in academic performance, resulting in a PMNP of 32.56% and a PMNP multiplier of 1.67.

5.3 Targeted pedagogical improvements

By enhancing the CHAID decision tree to uncover the main factors affecting the PMNP, it can effectively help teachers to propose targeted teaching improvement measures. For example, for node 13, even though the students' motivation for learning is interest, the PMNP in the university English (A) teaching mode, caused by poorer teaching tools, is 0.54. It can be seen that there is a significant difference between the PMNP multiplier (1.29) for node 13 and node 18. By comparing Node 13 and Node 18 it can be seen that in general, despite the important role of motivation, the impact of multimedia teaching tools on learning performance is more significant in the particular teaching mode. The reason for this is that students in the University English (A) mode of teaching have a lower foundation in English than those in the High Start and Happy English modes of teaching. As a result, these students have a greater reliance on computer-aided tools such as multimedia. Therefore, the use of multimedia teaching tools, such as designing high quality teaching PPT courseware, should be increased for these students.

6 Conclusions

This paper uses an enhanced CHAID decision tree based on genetic algorithm to uncover the main factors affecting PMNP, which can effectively help teachers achieve teaching optimisation. In the context of big data, this paper selects 1,805 students of non-English majors in a university as the respondents and conducts a study on the optimisation of university English teaching based on the enhanced decision tree model. According to the analysis results of the enhanced CHAID decision tree, the most important factors influencing the PMNP of the sample group are, in order, learning motivation, teaching means and teaching mode. Although this paper provides a more in-depth study of English teaching in universities, the subsequent guidance of student behaviour needs to be explored. In practical English language teaching, schools can base on the enhanced CHAID decision tree for reasonable statistics on teaching effectiveness. At the same time, the enhanced CHAID decision tree can help students make effective use of various resources around them in English teaching, and help develop their practical skills.

Acknowledgements

This research has been financed by the '14th Five-year Plan' project initiation in 2022 of the Chongqing Academy of Education Science 'Research on the Practice Model for Vocational Colleges Serving the Cultural Revitalization of Rural' (K22YG316323).

References

Abbasi, A. and Izadpanah, S. (2018) 'The relationship between critical thinking, its subscales and academic achievement of English language course: the predictability of educational success based on critical thinking. Predictability of educational success based on critical thinking', *Academy Journal of Educational Sciences*, Vol. 2, No. 2, pp.91–105.

Ahmadi, D. and Reza, M. (2018) 'The use of technology in English language learning: a literature review', *International Journal of Research in English Education*, Vol. 3, No. 2, pp.115–125.

Alsowat, H. (2016) 'An EFL flipped classroom teaching model: effects on English language higher-order thinking skills, student engagement and satisfaction', *Journal of Education and Practice*, Vol. 7, No. 9, pp.108–121.

Amerian, M. and Tajabadi, A. (2020) 'The role of culture in foreign language teaching textbooks: an evaluation of New Headway series from an intercultural perspective', *Intercultural Education*, Vol. 31, No. 6, pp.623–644.

Basal, A. (2015) 'The implementation of a flipped classroom in foreign language teaching', *Turkish Online Journal of Distance Education*, Vol. 16, No. 4, pp.28–37.

Baum, N. and Boughton, L. (2016) 'Public speaking: managing challenging people and situations', *The Journal of Medical Practice Management*, Vol. 31, No. 4, pp.251–253.

Candel-Mora, M.A. (2015) 'Attitudes towards intercultural communicative competence of English for specific purposes students', *Procedia-Social and Behavioral Sciences*, Vol. 178, pp.26–31.

Chen, L., Feng, G., Leong, C.W., et al. (2016) 'Designing an automated assessment of public speaking skills using multimodal cues', *Journal of Learning Analytics*, Vol. 3, No. 2, pp.261–281.

Chodkiewicz, J. and Miniszewska, J. (2015) 'The polish adaptation of the public speaking anxiety questionnaire', *Psychiatria Polska*, Vol. 49, No. 1, pp.95–105.

Chuang, H.H., Weng, C.Y. and Chen, C.H. (2018) 'Which students benefit most from a flipped classroom approach to language learning?', *British Journal of Educational British Journal of Educational Technology*, Vol. 49, No. 1, pp.56–68.

Doman, E. and Webb, M. (2016) 'The flipped experience for Chinese university students studying English as a foreign language', *Tesol Journal*, Vol. 8, No. 1, pp.102–141.

Eva, R. (2015) 'Techniques developing intercultural communicative competences in English language lessons', *Procedia-Social and Behavioral Sciences*, Vol. 186, No. 1, pp.939–943.

Freiermuth, M.R. (2017) ''I Found It!' A smartphone GPS treasure-hunting game in a flipped English class', *Innovation in Language Learning & Teaching*, Vol. 11, No. 11, pp.1–8.

Gerbensky, K.A. (2017) 'Creating a structured practice space with online mini-speeches', *Communication Teacher*, Vol. 31, No. 2, pp.70–73.

Handoyo, P.W. (2016) 'Engaging students in literature circles: vocational English reading progress', *Asia-Pacific Edu. Res.*, Vol. 25, No. 2, pp.347–359.

Hellwig, A.F.J. (2022) 'Multimodal, digital artefacts as learning tools in a university subject-specific English Language course', *Int. J. TESOL Studies*, Vol. 4, No. 2, pp.24–38.

Hsieh, J.S.C., Huang, Y.M. and Wu, W.C.V. (2016) 'Technological acceptance of LINE in flipped EFL oral training', *Computers in Human Behavior*, Vol. 70, No. 2, pp.178–190.

Hung, H.T. (2015) 'Flipping the classroom for English language learners to foster active learning', *Computer Assisted Language Learning*, Vol. 28, No. 1, pp.81–96.

Hung, H.T. (2017) 'Design-based research: redesign of an English language course using a flipped classroom approach', *Tesol Quarterly*, Vol. 51, No. 1, pp.180–192.

King, P.E. and Finn, A.N. (2017) 'A test of attention control theory in public speaking: cognitive load influences the relationship between state anxiety and verbal production', *Communication Education*, Vol. 66, No. 2, pp.168–182.

Komiya, R., Saitoh, T., Fuyuno, M., et al. (2017) 'Head pose estimation and motion analysis of public speaking videos', *International Journal of Software Innovation*, Vol. 5, No. 1, pp.57–71.

Kvashnina, O.S. and Martynko, E.A. (2016) 'Analyzing the potential of flipped classrooms in ESL teaching', *International Journal of Emerging Technologies in Learning*, Vol. 11, No. 3, pp.71–73.

Le, F.L., Le, F.L., Blackburn, K., et al. (2015) 'Student estimates of public speaking competency: the meaning extraction helper and video self-evaluation', *Communication Education*, Vol. 64, No. 3, pp.261–279.

Lee, B. (2017) 'Tell us ESP in a flipped classroom', *Eurasia Journal of Mathematics Science & Technology Education*, Vol. 13, No. 8, pp.4995–5007.

Lee, G. and Wallace, A. (2018) 'Flipped learning in the English as a foreign language classroom: outcomes and perceptions', *Tesol Quarterly*, Vol. 52, No. 1, pp.62–84.

Lertworaprachaya, Y., Yang, Y. and John, R. (2014) 'Interval-valued fuzzy decision trees with optimal neighbourhood perimeter', *Applied Soft Computing*, Vol. 2014, No. 24, pp.851–866.

Mancuso, C. and Miltenberger, R.G. (2016) 'Using habit reversal to decrease filled pauses in public speaking', *Journal of Applied Behavior Analysis*, Vol. 49, No. 1, pp.188–192.

Mari, H. and Mariko, A. (2015) 'ESL teacher advocacy beyond the classroom', *Teaching and Teacher Education*, Vol. 49, No. 1, pp.149–158.

Mashael, A.B. and Muna, A.R. (2016) 'Predicting student's final GPA using decision trees: a case study', *International Journal of Information and Education Technology*, Vol. 6, No. 7, pp.528–533.

Menegazzo, E. (2017) 'Flipped learning in LS and development of complex cognitive skill', *Lingue Antiche Emoderne*, Vol. 11, No. 6, pp.131–150.

Mercedes, I.C., Ridwan, M., Carmen-María, F.G., et al. (2019) 'Validating a model of effective teaching behaviour and student engagement: perspectives from Spanish students', *Learning Environments Research*, Vol. 2, No. 2, pp.471–493.

Perry, L.B. (2015) 'Improving lecture quality through training in public speaking', *Innovations in Education & Teaching Internatio.*, Vol. 52, No. 2, pp.207–217.

Ramos, L.S. (2015) 'The challenge of allowing everyone's voice to be heard: introducing civil dialogue in public speaking courses', *Communication Teacher*, Vol. 29, No. 1, pp.1–6.

Reinhardt, J. (2020) 'Metaphors for social media-enhanced foreign language teaching and learning', *Foreign Language Annals*, Vol. 53, No. 2, pp.234–242.

Şahin, S. (2020) 'Developing a checklist for English language teaching course book analysis', *International Journal of Education and Research*, Vol. 8, No. 1, pp.107–120.

Schneider, J., Börner, D., Van, R.P., et al. (2017) 'Presentation trainer: what experts and computers can tell about your nonverbal communication', *Journal of Computer Assisted Learning*, Vol. 33, No. 2, pp.164–177.

Spieler, C. and Miltenberger, R. (2017) 'Using awareness training to decrease nervous habits during public speaking', *Journal of Applied Behavior Analysis*, Vol. 50, No. 1, pp.38–47.

Stob, P. (2015) 'Talk like TED: the 9 public-speaking secrets of the world's top minds', *Quarterly Journal of Speech*, Vol. 101 No. 1, pp.311–314.

Suranakkharin, T. (2017) 'Using the flipped model to foster Thai learners' second language collocation knowledge', *The Southeast Asian Journal of English Language Studies*, Vol. 24, No. 3, pp.1–20.

Wang, B., Teo, T. and Yu, S.L. (2017) 'Teacher feedback to student oral presentations in EFL classrooms: a case study', *Journal of Education for Teaching: International Research and Pedagogy*, Vol. 43, No. 2, pp.262–264.

Ward, S. (2016) 'It's not the same thing: considering a path forward for teaching public speaking online', *Review of Communication*, Vol. 16, No. 2, pp.222–235.

Wickramarachchi, D.C., Robertson, B.L., Reale, M., et al. (2016) 'HHCART: an oblique decision tree', *Computational Statistics & Data Analysis*, Vol. 7, No. 96, pp.12–23.

Xu, X., Cai, H., Qiao, Z., et al. (2017) 'Impacts of park landscape structure on thermal environment using QuickBird and Landsat images', *Chinese Geographical Science*, Vol. 27, No. 5, pp.818–826.