Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A Study on Decision Tree Analysis Method of Teaching Quality Improvement for Teachers of Marketing in Higher Education Institutions

Huaying Yu[1,] ✉

[1] *Linyi Vocational College, Linyi, Shandong, 276000, China*

ABSTRACT

In order to explore the deficiencies in the teaching process of marketing majors in higher vocational colleges and further improve the teaching quality of marketing majors in higher vocational colleges. This paper utilizes the improved ID3 algorithm to construct the SLIQ data mining algorithm to improve the teaching quality of teachers of marketing majors in higher vocational colleges and universities. Using ID3 algorithm to build a decision tree to get the portraits of teachers and students, at the same time, in order to reduce the computational complexity of ID3 algorithm and the problem of multi-value bias, the concept of sample structure vector similarity is introduced, and the degree of information gain is optimized to get a more reasonable decision tree. On this basis, based on the improved ID3 data mining algorithm, a teaching quality assessment system for senior marketing majors based on SLIQ algorithm is designed, which identifies important factors affecting teachers' teaching quality by mining a large amount of data in the teaching process.The AUC value of the SLIQ data mining algorithm is 0.98, which can effectively improve the algorithm's generalization ability, and it has an excellent performance in the teaching quality assessment task. The performance is excellent. In this paper, we systematically identify “the principles of marketing” and “the degree of seriousness of teachers' homework correction” as the key factors to improve the teaching quality of marketing teachers. It provides a scientific basis for improving the quality of teachers' teaching.

Keywords: ID3 algorithm, SLIQ, information gain, data mining algorithm, teaching quality assessment

1. Introduction

In recent years, in order to strengthen the spiritual civilization construction, the state supports the development of education, especially the development of higher education is very fast, and the higher vocational is also expanding year by year, cultivating a large number of talents for the society and

✉ Corresponding author.
E-mail address: yhy20082020@163.com (H. Yu).
Received 30 May 2024; Revised 15 August 2024; Accepted 10 January 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-214

enterprises [1-2]. With the development of the economy, the society has an increasing demand for excellent marketing talents, but nowadays the marketing students trained in higher vocational education cannot meet the needs of enterprises in terms of quality [3-4]. The reason for this is that higher vocational marketing teaching is far away from the market and fails to keep up with the times [5].

In order to adapt to the changes in the demand for social talents, higher vocational colleges and universities have begun to pay attention to the construction of marketing specialties. On the one hand, vigorously improve the marketing training base, improve the marketing teaching hardware environment, on the other hand, strengthen the marketing faculty construction [6-7]. But marketing students are still difficult to get the favor of society and enterprises, many employers feedback really will do a good job in marketing staff is not a marketing professional students, marketing students in the workplace did not reflect their professional advantages, now marketing students employment encountered a relatively embarrassing scene [8-10]. Practice teaching, is the focus of modern teaching reform. With the saturation of theoretical talents in society, the demand of enterprises for practical skills talents with strong hands-on ability is also increasing. Strengthening practical teaching is the key work of cultivating technical skill talents [11-14]. At present, the reform of higher vocational, many general undergraduate colleges and universities began to be transformed into application-oriented colleges and universities. In the new period, how to improve the quality of practical teaching and cultivate excellent marketing professionals in applied colleges and universities is always a new idea of school running in applied colleges and universities [15-17].

Wang, H based on the fuzzy Delphi method on the higher vocational high marketing teaching to explore, uncovered the marketing teaching to be improved, and put forward targeted solution suggestions [18]. Chen, M. explored the marketing career teaching of the five also link specialization, as well as the teaching status quo, especially the insufficiencies and advantages of the integration of industry and education training mode, aimed at establishing and perfecting the market-adapted marketing personnel training system [19]. Chen, L et al. elaborated the characteristics of higher vocational teaching focusing on practice and students' skill development, of which off-campus practice and on-campus practical training are important components, and proposed to reconstruct the content of higher vocational teaching and use flipchart teaching materials as well as the project-based teaching methodology in order to improve the flexibility of the integrated course of network marketing [20]. The above research is mainly to explore the investigative research, discover the deficiencies of higher vocational marketing courses and deeply understand the current situation of higher vocational marketing teaching, and put forward some optimization suggestions from the level of talent training system, teaching methods and so on.

Based on the research of teaching technology mainly involves information technology as the core logic of the flipped can he, online revealed blended teaching mode, Yanbing, L to carry out controlled teaching experiments, revealing that the flipped classroom to a certain extent stimulate the enthusiasm of the marketing teaching classroom, obtaining a good marketing teaching effect [21]. Huang, Q et al. based on the theory of blended learning, around the higher vocational teaching Based on the blended learning theory, Huang, Q et al. designed a blended teaching model of marketing strategy around the three levels of learning scenario analysis, teaching design and teaching evaluation, which effectively enhanced students' motivation and practical ability in marketing course learning [22].

In this paper, we improved the ID3 algorithm and designed a teaching quality assessment system for marketing majors in higher vocational colleges based on SLIQ data mining algorithm. Firstly, the sample structure vector similarity was used to quantify the relationship between conditions and classification attributes, which solved the multi-value bias problem of ID3 algorithm. Subsequently, the smallest optimization information gain value is taken as the selection criterion, which realizes the simplification of the arithmetic process. The SLIQ algorithm in the teaching quality assessment system

utilizes the pre-sorting technique and hierarchical data storage structure to improve the operation efficiency. In addition, split point calculation method and DML pruning algorithm are introduced to improve the accuracy of algorithm classification. Finally, the comparison experiment verifies the performance of SLIQ algorithm and validates the effectiveness of the system of this paper in practical applications.

2. Decision tree analysis method

2.1. ID3 algorithm

The core idea of the ID3 algorithm [23] is to measure the selection of attributes in terms of information gain and select the attribute with the largest information gain after splitting. The algorithm uses top-down greedy search to traverse the possible decision space.The advantages of the ID3 algorithm include simplicity, interpretability, ability to handle categorical and continuous attributes, and some ability to handle missing values.

Let the set of training instances be S, $|S|=s$, divide the training set into m classes, denoted $C=\{S_1,S_2,\ldots,S_m\}$, let the number of training instances in class i be $|S_i|=C_i$, and denote the probability of a sample belonging to class i as $P(X_i)$, then there is $P(X_i)=C_i/s$. Then the degree of uncertainty (i.e., the entropy of information needed to make an accurate class judgment) of the decision tree for the division of C at this point is $H(S)$:

$$H(S)=-\sum_{i=1}^{t}P(S_i)\log_2 P(S_i) \tag{1}$$

If attribute A is chosen to divide the training set S, and let attribute A have attribute value $\{a_1,a_2,\ldots,a_n\}$, then the number of training instances of a_j belonging to class i is C_{ij}. Then, $P(S_i \mid A=a_j)=C_{ij}/S_i$, denotes the probability of belonging to class i when the value of test attribute A is a_j. Let the set of instances of $A=a_j$ be denoted as Y_j, then the conditional entropy of the training set for attribute A is equal to the uncertainty of the decision tree classification, i.e:

$$H(Y_j)=-\sum_{i=1}^{i}P(S_i \mid A=a_i)\log_2 P(S_i \mid A=a_j) \tag{2}$$

The information entropy of node A for categorization is derived from the subset divided by test attribute A:

$$H(S \mid A)=\sum_{j=1}^{i}P(A=a_j)H(Y_j) \tag{3}$$

Then the amount of information provided by attribute A for categorization, i.e., the information gain of attribute A, is denoted as $Gain(A)$:

$$Gain(A)=H(S)-H(S \mid A) \tag{4}$$

ID3 algorithm is mainly used to construct decision trees. It selects the best dividing attribute by calculating the information gain of each attribute, so as to construct a decision tree that can classify the training samples perfectly. In this paper, ID3 algorithm is used to construct a decision tree, analyze its leaf nodes in depth, and then complete the drawing of the portraits of marketing teachers and students. In the process of constructing the decision tree, the ID3 algorithm is found to have the problems of computational complexity and obvious multi-value bias, and the ID3 algorithm is improved by improving the above two problems.

2.2. Improved ID3 algorithm

Sample structure vector similarity is a numerical metric used to measure the spatial structure direction between attribute values of descriptive and categorical attributes in sample data. In positive space, a numerical value between 0 and 1 is calculated, and this number is the sample structure vector similarity value.

Where the greater the sample structure vector similarity describing the attribute, the more the attribute converges to the categorized sample attribute and vice versa. This value is used as a weighting factor so that it is involved in the selection of the descriptive attribute, thus optimizing the generation of the decision tree.

1) Structure similarity matrix. In order to get the value of the sample structure vector similarity, the first step is to construct the structure similarity matrix based on the sample, extract and calculate the values in the matrix, and then judge the degree of agreement between the description attributes and classification attributes in the sample data source when they are not used repeatedly.

Let there exist a sample training dataset S, $|S|=n$, dataset A is a descriptive attribute of training dataset S, attribute A takes values in the space of $A=[A_1,A_2,\ldots,A_p]$, A_i table the number of samples with different classifications of attribute A. The sequence of attribute A in the training data set S is $X=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x \end{bmatrix}$, and there exists data set B as a categorization attribute of the training data set S, which has m different values, then the value space of the categorization attribute B can be expressed as $B=[B_1,B_2,\ldots,B_q]$, and according to the number of samples under the same attribute value from the largest to the smallest to obtain the new $B'=[B_1',B_2',\ldots,B_q']$, and at the same time to obtain the sequence of attribute B' in the training data set as $Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y \end{bmatrix}$ a data sequence consisting of a descriptive original attribute A and a categorization attribute B together constitute the data sequence (X,Y) as:

$$(X,Y)=\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} \tag{5}$$

The description attribute A takes values in rows and the categorization attribute B takes values in columns, resulting in a matrix M with p rows and q columns:

$$M=\begin{bmatrix} a11 & a12 & \cdots & a1j & \cdots & a1q \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ ai1 & ai2 & \cdots & aij & \cdots & aiq \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ ap1 & ap2 & \cdots & pj & \cdots & apq \end{bmatrix} \tag{6}$$

a_{ij} denotes the number of samples of the data tuple in (X,Y) when the value of y is B': and the value of x is A_j. This structural similarity matrix describes the structural distribution of the attribute values between sequences X and Y in the sample data set.

By introducing the sample structural vector similarity, it is possible to quantify the relationship between the condition and the categorized attributes that could have been obtained from the dataset, to combine the distribution of attribute values of different condition attributes in the dataset with the inherent information gain of the ID3 algorithm, and thus to give a clear categorization definition of different condition attributes and to judge the categorization trend of the attribute. This ensures the importance of information gain in classification on the one hand, and circumvents the multi-value bias problem of ID3 algorithm on the other.

2) Vector similarity calculation. For two vectors X and Y of the same dimension in the space, the cosine of the angle between them:

$$Sim(X,Y)=\frac{|X\cdot Y|}{\|X\|\cdot\|Y\|} \tag{7}$$

It is called the vector similarity, which takes values in the range [0, 1] in positive space. Often $Sim(X,Y)$ is used to measure the similarity of a set of vectors. In the text, when $Sim(X,Y)$ is larger, it means that the convergence between vectors X and Y is better, but on the contrary, it is not easy to distinguish vectors X and Y; when $Sim(X,Y)$ is smaller, it means that the convergence between vectors X and Y is worse, i.e., the angle between vectors X and Y is larger, and the easier it is to distinguish vectors X and Y. $Sim(X,Y)$ can be used to assess the importance of the descriptive attribute to the overall database classification, $Sim(X,Y)$ the closer to 0, indicating that the correlation between X and Y is lower, indicating that the descriptive attribute plays a significant role in the data classification, which in turn indicates that the descriptive attribute is better able to reduce the overall information entropy, and therefore the descriptive attribute has greater information gain, and thus can become a classification node of the decision tree.

Through the construction of structural similarity matrix, the column vectors obtained by taking the values of description attribute and classification attribute are used to calculate the vector similarity under the description attribute, so that it can objectively reflect the degree of correlation between the description attribute and the classification attribute, and overcome the problem of multi-value bias that exists in the original ID3 algorithm, without being affected by the number of values taken.

3) Optimization of the information gain formula. In order to simplify the calculation, the set of training examples is divided into two categories of positive and negative examples, and let the number of positive and negative example sets be p and n, respectively, and the value of attribute A is $\{a_1,a_2,\ldots,a_n\}$. When there is $A=a_i$, there are a total of n_i+p_i elements in the category to which they belong, i.e., p_i positive examples and n_i negative examples. According to the principle of ID3 algorithm is obtained from the information gain calculation formula:

$$Gain(A)=-\sum_{i=1}^{c}p_i\log_2(p_i)-\sum_{j=1}^{n}\frac{n_j+p_j}{n+p}\times\sum_{i=1}^{m}p_{ij}\log_2(p_{ij}) \tag{8}$$

is certain for each attribute $H(S)$, so comparing the conditional entropy $H(S|A)$ of attribute A, i.e:

$$H(S|A)=\frac{1}{(n+p)\ln 2}\sum_{j=1}^{n}(-n_j\ln\frac{n_j}{n_j+p_j}-p_j\ln\frac{p_j}{n_j+p_j}) \tag{9}$$

According to the McLaughlin series expansion of the function $\ln(1+x)$, taking the second order approximation when $x\in(-1,1)$ has:

$$\ln(1+x) \approx x - \frac{1}{2}x^2 \tag{10}$$

Since $0 < p_j(n_j + p_j) < 1$ and $0 < n_j / (n_j + p_j) < 1$, and $1/(n+p)\ln 2$ are certain values and can be ignored, further simplification of the conditional entropy $H(S \mid A)$ yields $H'(S \mid A)$:

$$\begin{aligned} H(S \mid A) = \sum_{j=1}^{n} -p_j \left[\left(-\frac{n_j}{n_j + p_j} \right) - \frac{1}{2} \left(-\frac{n_j}{n_j + p_j} \right)^2 \right] \\ -n_j \left[\left(-\frac{p_j}{n_j + p_j} \right) - \frac{1}{2} \left(-\frac{p_j}{n_j + p_j} \right)^2 \right] = \sum_{j=1}^{n} \frac{5 p_j n_j}{2(p_j + n_j)} \end{aligned} \tag{11}$$

Since $Gain(A) = H(S) - H(S \mid A)$, where $H(S)$ is the desired entropy for dividing the sample dataset, is a certain value, it is sufficient to compare the information entropy describing the attributes, i.e., to compare the simplified conditional entropy $H'(S \mid A)$

$$H'(S \mid A) = \sum_{j=1}^{n} \frac{p_j n_j}{p_j + n_j} \tag{12}$$

The simplified conditional entropy, which is also the minimum value of $H'(S \mid A)$, is chosen as the classification attribute node.

The ID3 algorithm based on the similarity of sample structure vectors mainly consists of two modifications: the improvement of the ID3 algorithm and the simplification of the formula of the ID3 algorithm. By combining the two, the optimized information gain of ID3 algorithm based on sample structure vector similarity is obtained $Gain'(A)$:

$$Gain'(A) = Sim(X, Y) \times H'(S \mid A) \tag{13}$$

$Gain'(A)$ is used as the new attribute selection criterion, i.e., the smallest optimized information gain value is selected. $Gain'(A)$ avoids logarithmic operations on the basis of $Gain(A)$ and has only basic operations, so the algorithm is simple to implement and much faster than the original. At the same time, $Gain'(A)$ introduces objective weighting coefficients that can also overcome the problem of multi-value bias. If the result of the calculation appears to have the same conditional entropy value of the attribute, the inherent information gain of the attribute is calculated at this time.

3. Data mining based teaching quality assessment system for teachers

3.1. Structure of the Teacher Teaching Quality Assessment System

In the data mining-based teaching quality assessment system for marketing majors in higher vocational colleges, the structural framework of the designed teaching quality assessment system for teachers is shown in Fig. 1. The structure adopts a layered architecture approach and consists of a representation layer, a processing layer and a mining layer. The representation layer is responsible for providing user interaction interface and storing user information; the processing layer handles user requests and calls the corresponding data mining module to execute them: the mining layer is the core of the system, which contains several groups of data mining components, each of which implements a data mining algorithm. The designed structural framework of the data mining system can improve the processing efficiency of the system and the accuracy of data mining, and provide scientific and effective technical support for the assessment of teaching quality of marketing majors in higher vocational colleges.

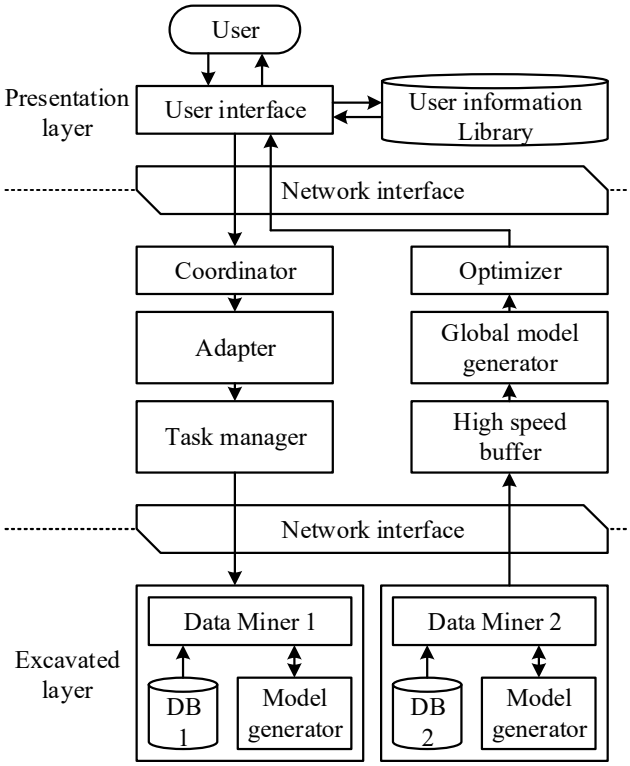

Fig. 1. Data mining system structure

3.2. SLIQ data mining algorithm with improved ID3 algorithm

3.2.1. SLIQ data mining process. SLIQ data mining algorithm [24] is a decision tree construction algorithm for large-scale datasets. The algorithm significantly improves the efficiency and scalability of data processing by introducing pre-sorting techniques and hierarchical data storage structure. Unlike traditional algorithms that sort at each node, SLIQ sorts the data only once in the preprocessing stage, reducing the computational complexity. In addition, the algorithm uses split-point selection and pruning mechanisms to optimize the decision tree generation process and improve the accuracy of classification. In the analysis of educational data applications, especially in the assessment of teaching quality of marketing majors in higher vocational colleges and universities, the SLIQ algorithm can efficiently process and analyze the data to provide a scientific basis for teaching decisions.

SLIQ algorithm first introduces 2 data structures, attribute list and class histogram, in order to find the best split point of each attribute quickly and accurately.The flow of SLIQ algorithm is shown in Fig. 2. First, all records are pre-sorted by attribute values to eliminate the need for sorting at each node of the decision tree. Second, the SLIQ algorithm uses a breadth-first search strategy to split all leaf nodes in the decision tree at the same time to ensure that each split is based on the current optimal split criterion. During the construction of the decision tree, the class histogram is used to store the class distribution of each node, which is used to calculate the optimal split point. At the same time, the algorithm adopts the minimum description length (MDL) pruning method, which avoids overfitting and ensures the generalization ability of the model. Aiming at the characteristics of teaching quality assessment of marketing majors in higher vocational colleges, the process improves the efficiency and accuracy of large-scale teaching data processing.

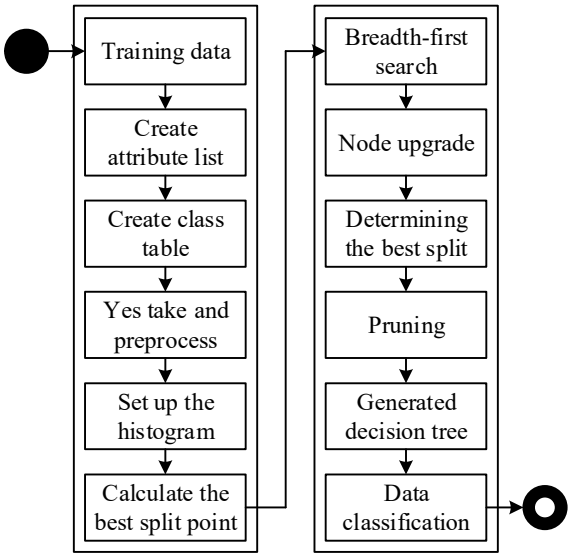

Fig. 2. SLIQ algorithm flow

3.2.2. Calculating the optimal splitting point. The SLIQ algorithm determines the optimal splitting point by evaluating the information gain [25] or the Gini index [26] to optimize the splitting of nodes when constructing the decision tree. Specifically, the information gain is calculated based on the reduction in the uncertainty of the dataset before and after attribute splitting. For a given dataset D, its uncertainty can be represented by the entropy $H(D)$, which is calculated as:

$$H(D) = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{14}$$

where m is the number of categories and p_i is the relative frequency of the i rd category in dataset D.

When selecting split attributes and split points, the algorithm calculates the information gain from each possible split and selects the split point with the largest information gain as the best split point. The information gain $IG(D, A)$ is calculated as:

$$IG(D, A) = H(D) - \sum_{a \in Valars(A)} \frac{|D_e|}{|D|} H(D_e) \tag{15}$$

where A is the candidate split attribute, $Values(A)$ is all possible values of attribute A, D_e is the subset of attribute A when its value is v, $|D|$ is the size of dataset D, and $|D_e|$ is the size of dataset D.

The Gini index is another method of assessing the quality of splits and is used to measure the purity of the dataset. The smaller the Gini index of a node, the higher the purity of the dataset. The Gini index $Gini(D)$ is calculated by the formula:

$$Gini(D) = 1 - \sum_{i=1}^{m} P_i^2 \tag{16}$$

For each attribute, the SLIQ algorithm calculates the weighted Gini index of the split subset and selects the split point that minimizes the weighted Gini index as the best split point.

This process allows the system to accurately identify the key factors affecting the quality of teaching and learning in a teaching quality assessment system for marketing majors in higher vocational colleges and universities.By analyzing the students' learning data, such as grades, participation, and

feedbacks, the proposed assessment system is able to give effective paths to improve the quality of teaching and learning.

3.2.3. DML Pruning Algorithm. The algorithm designed in this paper uses decision tree pruning technique to improve the generalization ability by minimizing the complexity of the decision tree, thus avoiding the overfitting phenomenon.The DML pruning algorithm is based on an information theoretic principle that the optimal model for a given dataset is the one that is able to describe the data in terms of the shortest description length (i.e., the minimum amount of information).The DML pruning process can be expressed as an optimization problem whose optimization objective is to make the cost is minimized.

$$Cost(T) = -\log L(T \mid D) + \frac{1}{2} k \log N \tag{17}$$

where T denotes the decision tree model, $L(T \mid D)$ is the likelihood function for a given dataset D in model T, k is the number of parameters of the model (which in decision trees usually corresponds to the number of leaf nodes in the tree), and N is the number of samples in dataset D. $-\log L(T \mid D)$ is the negative log-likelihood of how well the model T fits the given data set D, and $\frac{1}{2} k \log N$ is a penalty term for model complexity, measured by the size of the tree.

The application of the DML pruning algorithm in the Teaching Quality Assessment System for Marketing Programs in Higher Education Institutions ensures that the decision tree model maintains a large enough fit without losing the ability to predict unknown data due to the over-complexity of the model. By pruning the decision tree, the system is able to eliminate attributes that do not contribute much to the assessment of teaching quality, thus simplifying the assessment model and improving the efficiency and accuracy of the assessment. This process is crucial for identifying and reinforcing the key factors affecting the quality of English teaching in higher education institutions, and helps educational administrators and teachers develop more scientific and effective teaching re-training strategies based on data-driven insights.

4. Experimentation and Analysis of Teachers' Teaching Quality Improvement

4.1. Experiments on performance testing of the teaching quality assessment system

Two hundred students majoring in marketing in a higher vocational college are selected as experimental subjects to test the performance of the proposed system for distance teaching quality assessment. System testing is a key step to ensure the quality of the system. Before putting the system into use, it is necessary to determine the quality of the system software, test the performance of the system, and discover the problems in the hardware and software. Componentization technology was used for this experiment.

After 200 students of the marketing program of the university and three teachers (marketing teacher, advertising teacher, service marketing teacher) conducted distance teaching, the results of 200 students' assessment of teaching quality using the system of this paper are shown in Table 1. The experimental results show that the system of this paper can accurately assess the distance teaching quality of different teachers, and the assessment indexes of the system of this paper are extremely comprehensive. Comprehensive scoring standards of 200 students, it can be seen that the teaching quality assessment results of marketing teachers, advertising teachers, service marketing teachers are 80.69 points, 81.39 points, 84.44 points, the quality of distance teaching of the three teachers are at an excellent level, but there is still room for progress, the teacher can be adjusted to the individual indicators through the assessment results at the right time, to improve the quality of distance teaching.

Teachers can make timely adjustments to individual indicators through the assessment results to improve the quality of distance teaching.

Table 1. The results of the remote teaching quality assessment of this system

Evaluation index	Marketing teacher	Advertising teacher	Service marketing teacher
Teaching content	76.59	83.47	84.65
Teacher teaching level	81.44	87.51	74.60
Teaching richness	83.59	81.49	86.52
Teaching management function	78.46	82.45	89.68
Real-time information of teaching data	81.28	84.65	91.55
Navigation design	77.00	86.76	84.16
Retrieval design	74.24	82.44	85.44
Online test function	77.34	76.89	82.68
Student interaction	74.19	74.85	94.54
Compatibility	71.39	77.63	74.23
Operating maintenance cost	72.63	71.63	85.48
Safety stability	81.67	82.64	87.59
User rate	83.64	84.56	86.55
Student coaching	87.42	74.8	84.22
System development cost	86.49	82.62	83.41
Course learning rate	89.42	83.61	79.82
Normality	87.15	81.64	84.54
Generality	89.80	86.94	83.30
Progressive property	79.43	79.88	81.36
Average	80.69	81.39	84.44

To ensure the accuracy of the experimental results, 200 students and service marketing instructors were scheduled to repeat the simulation experiment six times. The experiments were completed in a real simulation environment to test the functional and operational aspects of the system software. In order to visualize the evaluation performance of this paper's system, this paper's system is compared with the hierarchical analysis system and the fuzzy evaluation system, and the results of the connection speed comparison of 10 simulation experiments of the three systems in the campus network are shown in Figure 3. The connection speed comparison results of 10 simulation experiments of the three systems in the case of extranet are shown in Fig. 4. The results of the connection speed comparison of different systems for 10 simulation experiments under extranet situation with 200 students accessing the three systems at the same time are shown in Figure 5. Outside the network situation, the network peak 200 students simultaneously access three systems, 10 simulation experiments different system connection speed comparison results shown in Figure 6.

From the experimental results in Fig. 3 to Fig. 6, it can be seen that the connection time of the hierarchical analysis system and the fuzzy assessment system for distance teaching quality assessment in higher vocational colleges and universities is higher than 30 ms, and the connection time of the proposed system for distance teaching quality assessment in higher vocational colleges and universities in this paper is between 8 and 15 ms. It is concluded that the connection speed of the proposed system is qualified, indicating that the system in this paper meets the design requirements.

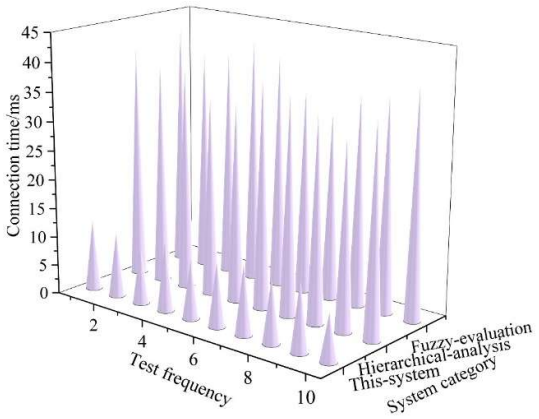

Fig. 3. Comparison of link speed in campus network

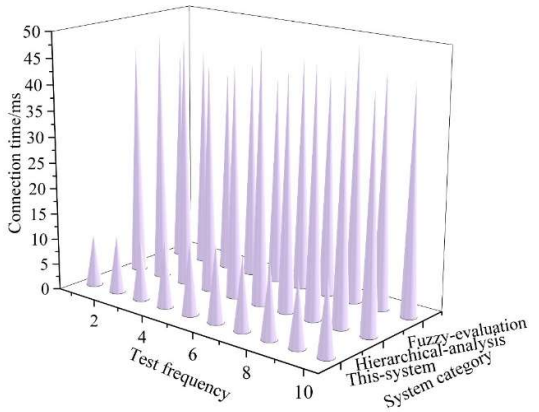

Fig. 4. Comparison of connection speed in external network

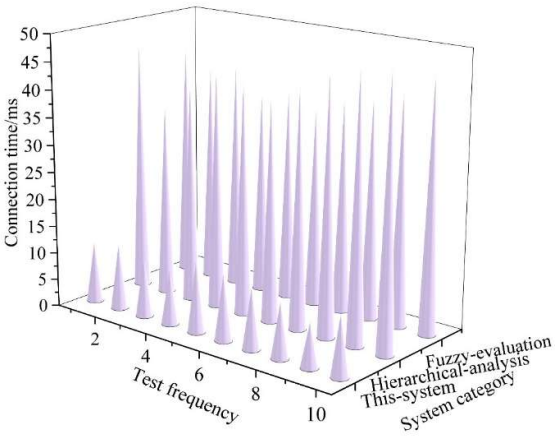

Fig. 5. Compare the connection speed of 200 users

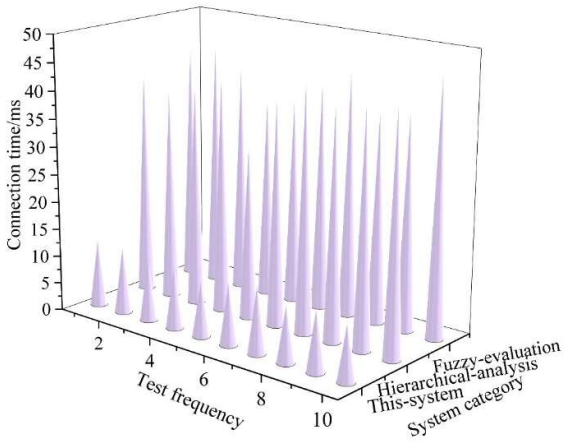

Fig. 6. Network peak speed contrast

4.2. Results of the evaluation of student performance

The experimental data in this chapter comes from the student grades of four classes in the marketing department of a higher education institution in the marketing program in the class of 2019. The above proposed system is used to model the students' pre-grade data and predict the average grade point average of the major in the later period. The importance of each course is obtained from the prediction results. As a result, teachers can focus on the teaching process to improve the quality of teaching and achieve the purpose of training talents.

The data of student performance evaluation comes from 8758 results of 200 students in a 2019 marketing college. Since the collected data are incomplete redundant data containing noise, the data need to be preprocessed. There are many number of attributes in the raw data, and some irrelevant attributes such as academic year, credits, class, and courseability are removed. The student achievement data is discretized using split-box method. The processed data will be categorized into five grades as follows: below 60 is failing, 60-70 is passing, 70-80 is moderate, 80-90 is good and 90-100 is excellent.

This experiment is implemented using the python language platform Anaconda3. The structure of the student performance evaluation model was first constructed. First, two important parameters in the model are determined: the value of the number of variables mtry for nodes and the number of trees ntree. It is obtained through experiments that when the number of decision trees takes a value greater than 350, the error rate tends to stabilize, and in this way the ntee value is set to 350. From the experiment it is obtained that when the number of variables selected for the decision tree nodes is 3, the mean value of the model's misclassification rate is the lowest, and the experimental results are shown in Fig. 7.

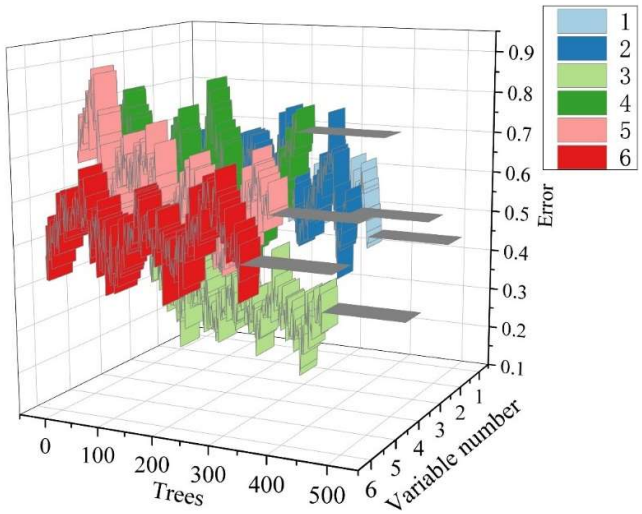

Fig. 7. Error diagram of decision tree quantity and model error

Students' grades from the first to fourth semesters were used to predict their grades in the fifth semester of the major and to rank the courses that would affect the next semester. The course importance ratio is shown in Figure 8.

Among these courses, "Principles of Marketing" has the greatest impact on students' professional learning courses. This is followed by "Microeconomics", "Macroeconomics", "Principles of Management", "Statistics", "Data Analysis", and "Accounting and Financial Management". "Advertising" and "foreign marketing" have little impact on students' performance.

According to the results of the comparison of the ordering of the important procedures of the two independent variables in the model obtained from the experiment, the practical class results have less impact on the later students' professional learning, and in the future teaching process, it can be targeted to have a tendency to teach students, and lay a good foundation for the students' learning of the subsequent courses.

In this study, when processing the data on students' grades, the model obtained after discretization is not particularly satisfactory due to the problems of missing data and multiple make-up test values in the collected data. The effect of other factors on grades was also not given much consideration. In future research, other factors can be considered and compared with multiple models to get more accurate results.

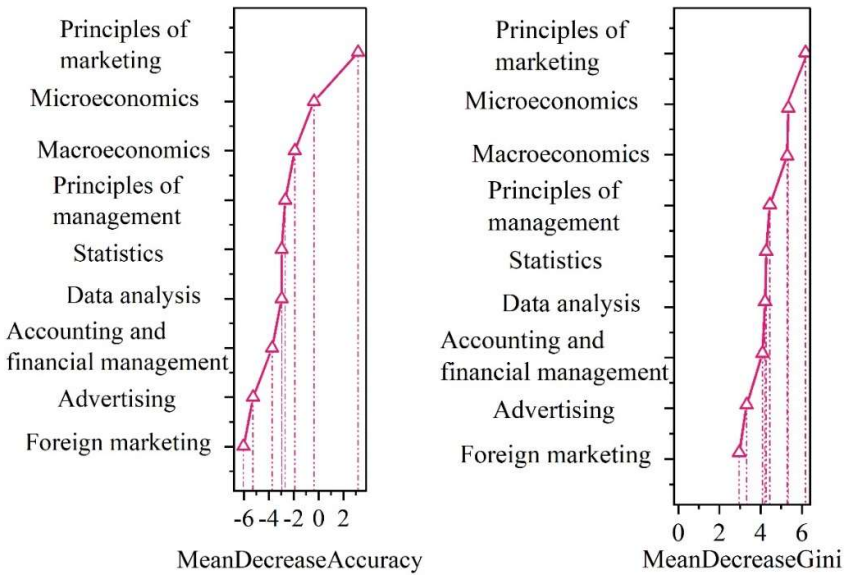

Fig. 8. The degree of the course is significantly compared

4.3. Analysis of factors affecting the assessment of teaching quality

For the data mining system in this paper, the most concerned about the model before entering the model is whether the model is overfitting, student data is more chaotic and numerous, deep learning type of model is more likely to have overfitting phenomenon in the training process, in order to address this issue need to be used in the training process to validate the collection of cross-validation methods, and continuous training, to see whether the model's accuracy can be sustained to improve and meet expectations. Figure 9 shows the model cross-validation training fluctuation chart, you can see that the constant cross-validation in the accuracy of the model does have an impact, but the accuracy rate to reach a smooth state when the data can still be maintained in the vicinity of 0.93, in line with the expectations of the model entry.

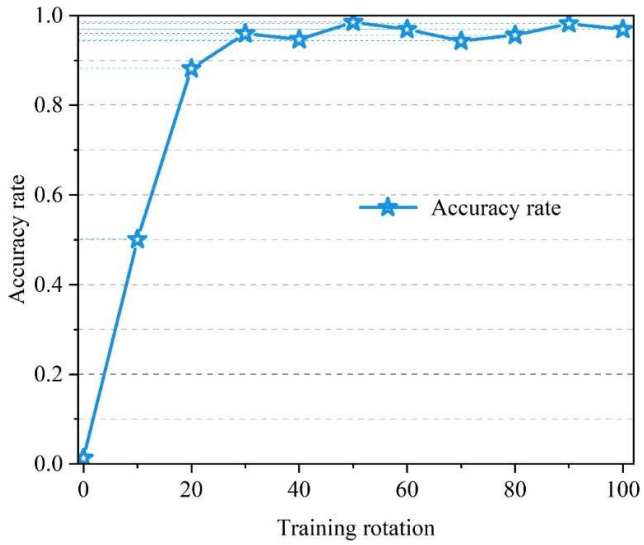

Fig. 9. Model cross-verification training accuracy variation diagram

The effect of data mining of teaching quality assessment model is good or bad to determine the analysis of teaching quality assessment, the two types of models of teacher and student data and teaching data are used to integrate to get a comprehensive teaching quality assessment model, and C4.5, XGBoost, CART, SLIQ,4 data mining techniques are selected as the method of model data mining. The ROC curve of the four algorithms is analyzed by experimental comparison as shown in Figure 10, the ROC curve of this paper's model is above the algorithm curves of C4.5 and XGBoost, which is closer to the point (0,1), and it is obvious that the AUC value of this paper's model using the SLIQ method is also the largest, so the model of this paper's model of the assessment model of the quality of teaching and learning is better in terms of generalization performance.

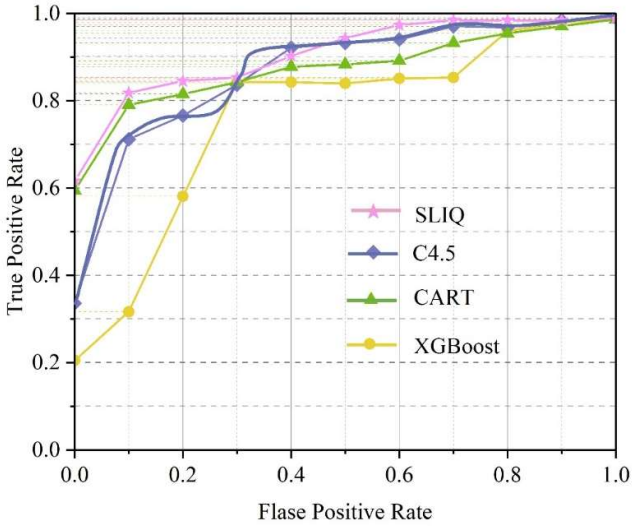

Fig. 10. Roc curve of four algorithms

Experiments were conducted for each of the 4 algorithms, and the accuracy, recall, P-R reconciliation mean and AUC values for the 4 algorithms are shown in Table 2. The accuracy and recall of the 4 fusion algorithms are analyzed: the C4.5 model has a higher accuracy rate, but the recall performance is poor and the model performance is not balanced. XGBoost is not effective, although the recall is a bit better than the C4.5 model, but the accuracy rate is too low and the prediction precision is not high. CART is better than the first two methods, but CART is worse in terms of efficiency.SLIQ model performs well in both recall and accuracy, which indicates that the prediction accuracy of the model in this paper is guaranteed under the premise of the model also guarantees the number of correct predictions, and the F1 value is also very high, which indicates that the advantages of this data mining technique are greater compared to the traditional methods.The SLIQ method, as the final data mining technique makes the AUC of the model also reach a high value of 0.98, which indicates that the overall generalization performance of the model is also optimal. In conclusion, SLIQ data mining algorithm as a teaching quality assessment model can effectively analyze the influencing factors of teaching quality.

Table 2. Four algorithms are compared

Algorithm	P	R	F1	AUC
C4.5	0.88	0.71	0.65	0.93
XGBoost	0.65	0.73	0.67	0.87
CART	0.84	0.84	0.78	0.92
SLIQ	0.88	0.89	0.85	0.98

The data used in this experiment are the teaching data of a higher vocational college in the year 2020-2021 and the related student data and teacher data, in order to protect personal information the experimental data is the real data after desensitization. The original feature influencing factors in the trained teaching quality assessment model are analyzed, and the distribution of influence weights is statistically shown in Figure 11. Through Figure 11, it can be seen that most of the features of the degree of influence is very small, and the degree of influence between 0.15-0.25 appeared between the peak of the secondary fluctuations, which can be seen that the high-frequency effective features should be in the middle of this, from the number of features in the degree of influence of the features of the analysis of the curve is smoother, it can be seen that the degree of influence of the features of the features of the degree of influence of the more uniform, the curve performance of the smooth and steady, the model features of the degree of influence of the analysis is also more Reliable, so through

this model to select the higher impact impression factor features to view can be derived from the higher vocational colleges and universities in the near future to improve the quality of teaching and learning of powerful factors.

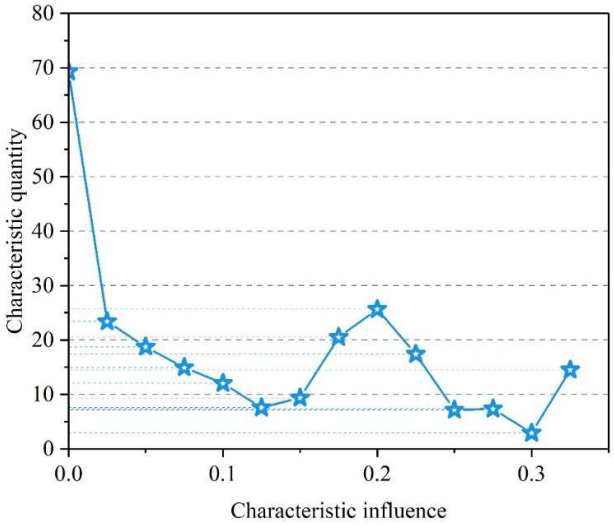

Fig. 11. Characteristics influence degree and quantity analysis

In this experiment, the Top10 influencing factors with higher degree of influence and the degree of influence are selected and plotted as shown in Figure 12. It can be found that most of the characteristics with high impact still come from the teaching factors related to the teaching skills, the reasonable degree of content, the number of class transfers and the classroom atmosphere have a greater impact on teaching quality, in which the degree of seriousness of the teacher's homework correction has the greatest impact, 0.34, which is also in the teaching of marketing majors to improve the teaching of some of the faster ways and means can be improved from the enhancement of the teaching skills, the reasonable degree of content, The number of transferring classes and classroom atmosphere and other aspects of teaching improvement, so as to improve the teaching quality of marketing professional teachers in higher vocational colleges and universities.

At the same time, the qualification level of marketing teachers will also have a certain impact on the teaching quality of marketing courses, the professionalism of teachers and excellent expert title from the side will also give students a certain impact of quality teachers, which will motivate students to study harder. The student side is also the higher impact of the student's enrollment results, which shows that the impact of a good learning foundation on the quality of teaching is also more prominent.

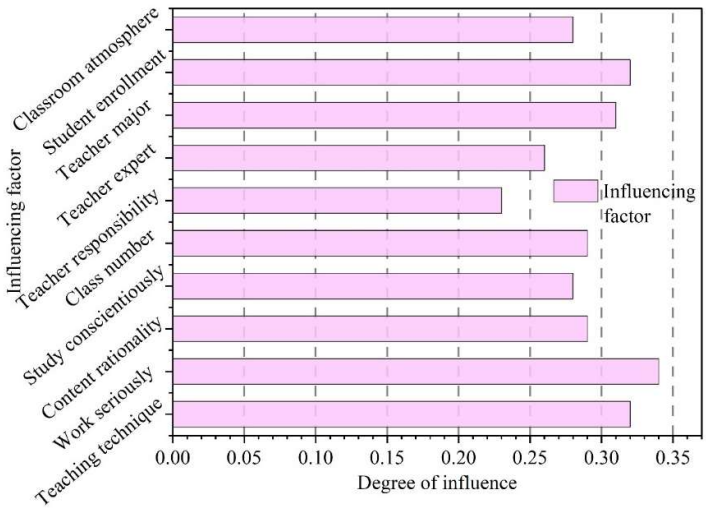

Fig. 12. Top10 influence factors and influence

5. Conclusion

In this paper, the factors affecting the teaching quality of marketing majors in higher vocational colleges and universities are quickly located and scientifically evaluated through data mining algorithms, so as to better extract the important factors affecting the teaching quality for the school.

1) The constructed decision tree algorithm effectively reduces the computational complexity and improves the generalization ability of the algorithm by minimizing the complexity of the decision tree, and compared with the three comparative algorithms of C4.5, XGBoost, and CART, the algorithm of this paper has the largest AUC value of 0.98. And the connection speed of the teaching quality assessment system based on the data mining algorithm of this paper is better than that of the hierarchical analysis system and fuzzy assessment system, which has better performance in the task of teaching quality improvement.

2) The teaching quality assessment system for marketing majors in higher vocational colleges based on the decision tree analysis algorithm demonstrates the great potential of the application of data mining technology in the field of education, and provides a brand-new tool for the assessment and enhancement of education quality. By precisely analyzing the teaching data of marketing majors, the key factors affecting teaching quality are found, and the “Principles of Marketing” course has the greatest influence on students' learning courses. Teachers' influencing factors have a greater impact on teaching factors, including teaching skills, reasonable content, the number of class transfers and the classroom atmosphere have a greater impact on the quality of teaching, and the degree of seriousness of the teacher's homework correction has the greatest impact. The mining of the above influencing factors provides a direction to promote the improvement of teaching quality.

References

[1] Rafdinal, W., Mulyawan, I., Juniarti, C., & Asrilsyak, S. (2020). The Decision of Prospective Students to Choose A Vocational College: The Role of the Marketing Mix and Image. Sriwijaya International Journal of Dynamic Economics and Business, 279-288.

[2] Kang, X. (2023). Application of Undergraduate Vocational Education Reform in the Digital Transformation of Marketing. Journal of Education and Educational Research, 5(1), 107-111.

[3] Qian, Z., & Xie, J. (2021). The effect of new media marketing integrated with professional course teaching in higher vocational colleges on alleviating college students’anxiety disorder. Psychiatria danubina, 33(suppl 6), 94-96.

[4] Barkas, L. A., Scott, J. M., Hadley, K., & Dixon-Todd, Y. (2021). Marketing students' meta-skills and employability: between the lines of social capital in the context of the teaching excellence framework. Education+ Training, 63(4), 545-561.

[5] John, S. P., & De Villiers, R. (2024). Factors affecting the success of marketing in higher education: a relationship marketing perspective. Journal of Marketing for Higher Education, 34(2), 875-894.

[6] Camilleri, M. A. (2019). Higher education marketing: Opportunities and challenges in the digital era. Camilleri, MA (2019). Higher Education Marketing: Opportunities and Challenges in the Digital Era. Academia, 0 (16-17), 4-28.

[7] Thi Kim Le, T., & Thi Tran, B. (2021, July). Student Satisfaction of the Blended Learning Implementation in Email Marketing Course: A Case Study at a Vocational College in Vietnam. In Proceedings of the 5th International Conference on Education and Multimedia Technology (pp. 44-50).

[8] Samaddar, K., & Menon, P. (2019). Marketing Higher Education: A Study on Attaining Sustainability through Redesigning Curriculum. SAMVAD, 18, 84-93.

[9] Ruilin, Z., & Lin, L. (2023). Research on the integration mode of production and teaching of digital marketing talents training in application-oriented universities. International Journal of New Developments in Education, 5(17).

[10] Angulo-Ruiz, F., Pergelova, A., & Cheben, J. (2016). The relevance of marketing activities for higher education institutions. International marketing of higher education, 13-45.

[11] Jin, Y., Wang, Q., Wang, Y., Geng, K., Ren, Y., & Wei, Y. (2020). Talent Training Mode of Big Data Marketing Major in Higher Vocational Colleges Based on: Scientific Research Assistance, Competition Driven. Education and Training Combination, 3(3).

[12] Sha, Y. (2018). Discussion on the Methods and Strategies of Personalized Training for Marketing Majors in Higher Vocational Colleges. Executive Chairman.

[13] Li, Z. (2022). Construction of Marketing Curriculum System Based on Blending Learning “3+ 2” Joint Training of Higher Vocational and Undergraduate Education Using NLP for Marketing Document Management and Information Retrieval. ACM Transactions on Asian and Low-Resource Language Information Processing, 21(6), 1-21.

[14] Zhu, Q. (2022). TEACHING PRACTICE REFORM OF FILM AND TELEVISION MAJOR IN HIGHER VOCATIONAL COLLEGES UNDER THE GUIDANCE OF SHORT VIDEO MARKETING. Psychiatria Danubina, 34(suppl 1), 299-300.

[15] Luthfi, B. R., & Jatmika, S. (2024). Marketing Industry Class Management in Vocational High Schools. Didaktika: Jurnal Kependidikan, 13(001 Des), 783-796.

[16] Gordillo, L. D., Domínguez, B. M., Vega, C., De la Cruz, A., & Angeles, M. (2020). Educational Marketing as a Strategy for the Satisfaction of University Students. Journal of Educational Psychology-Propositos y Representaciones, 8.

[17] Li, F., Liu, F., & Hu, X. (2018, December). Research on the Improvement Strategy of Marketing Talents Training in Secondary Vocational Schools. In 2018 2nd International Conference on Education Innovation and Social Science (ICEISS 2018) (pp. 17-20). Atlantis Press.

[18] Wang, H. (2023). Teaching Evaluation of Marketing Specialty in Higher Vocational Colleges Based on Fuzzy Delphi Method. marketing, 6(20), 93-99.

[19] Chen, M. (2022, July). The Application of Information Technology in Classified Training and Layered Teaching of Marketing Specialty in Higher Vocational Colleges Under the Integration of Industry and

Education. In International Conference on Frontier Computing (pp. 1106-1115). Singapore: Springer Nature Singapore.

[20] Chen, L., Lin, X., & Pan, Y. (2023). Exploration of Innovative Training Mode of School-enterprise Network Marketing: Take the" Volcano" Training base of Wenzhou Polytechnic as an example. International Journal of Education and Humanities, 8(2), 86-89.

[21] Yanbing, L. (2019). Evaluating instructional effects of flipped classroom in higher vocational colleges: a case study on marketing practice course. International Journal of Information and Education Technology, 9(4), 274-279.

[22] Huang, Q., Li, X., & Liu, H. (2022, July). Blended Learning Design for" Marketing Planning" Course in Higher Vocational Education. In Proceedings of the 6th International Conference on Education and Multimedia Technology (pp. 13-19).

[23] Huang Ke & Wang Tingxuan. (2024). Optimized Application of the Decision Tree ID3 Algorithm Based on Big Data in Sports Performance Management. International Journal of e-Collaboration (IJeC)(1),1-20.

[24] Narasimha Prasad L V & Mannava Munirathnam Naidu. (2014). CC-SLIQ: Performance Enhancement with 2k Split Points in SLIQ Decision Tree Algorithm. IAENG Internaitonal journal of computer science(3),163-173.

[25] Daniel Lichte. (2025). Combining Bayesian Networks and MCDA methods to maximise information gain during reconnaissance in emergency situations. Journal of Safety Science and Resilience(1),38-47.

[26] Guofeng Jin,Anbo Ming & Wei Zhang. (2024). Mathematical Mechanism of Gini Index Used for Multiple-Impulse Phenomenon Characterization. Aerospace(12),1034-1034.