Research Article

Interactive Teaching System for Remote Vocal Singing Based on Decision Tree Algorithm

Xiaotao Wang and Xiaoliang Liu

1 Jincheng Institute of Technology, Jincheng 048000, Shanxi, China
2 Weifang University of Science and Technology School of Sergeancy, Shouguang 262700, Shandong, China

Correspondence should be addressed to Xiaoliang Liu; liuxiaoliang@wfust.edu.cn

Received 6 April 2022; Revised 18 May 2022; Accepted 25 May 2022; Published 29 June 2022

As the global pandemic rapidly spreads, distance learning is emerging as a new method of instruction. Particularly, traditional courses that must be taught in-person, such as vocal singing instruction, require immediate adaptation to the COVID-19 and the new distance learning model. However, in the process of distance learning, it is frequently impossible to tailor and personalize instruction, particularly for vocal singing courses. Educational researchers are confronted with the pressing issue of how to extract useful and personalized patterns from a large volume of learner data in order to customize and individualize instruction. In this paper, we propose applying the decision tree method from data mining technology to a vocal singing education system by categorizing students according to the model. Once the characteristics of various learners have been stored in the corresponding user database, teachers can access timely information regarding the learners’ most recent learning situation. This can be used as the basis for differentiating instructional strategies for various learners. This allows the instructor to designate individualized teaching and learning strategies for each student.

1. Introduction

Education has always been at the heart of the work of colleges and universities, and enhancing the academic performance of students is one of the teaching goals of each institution. On the one hand, students’ performance reflects their knowledge mastery, and on the other, the assessment of students’ performance is also an essential criterion for determining the quality of instruction and the teaching level of teachers. Nonetheless, as China’s education industry develops, many universities will inevitably increase their enrollment. The increase in the number of students makes it difficult for teachers to fully comprehend the learning situation of each student in the classroom, so the analysis of student performance becomes the primary means by which teachers comprehend the learning situation of students.

Due to the national support for education informatization, most schools in China have established academic management systems [1]. In 2006, our university also developed the first version of the Academic Star Undergraduate Teaching System, which helps students and teachers to handle their affairs [2]. A large amount of data is stored in the system, including students’ personal information, grades of all courses during their study period, and students’ grade 4 and 6 results. However, currently, these pieces of exam-related information are simply stored in the system’s database, and if these grade data are analyzed, they can often yield a lot of useful information for optimizing teaching methods. Traditionally, the analysis of students’ performance data is still limited to simple querying of results, counting the number of students in each score range, the mean, variance, median of the results, and so on. Besides these pieces of information, a lot of imperceptible information is lost in the huge amount of data, for example, whether students with good scores in advanced mathematics have good scores in other courses and which leading courses have the greatest impact on the course after a subsequent course is completed. In the actual teaching process, there are
a lot of similar problems, so we can analyze a large number of students’ performance to find the basic courses that have a greater impact on students’ performance in a certain specialized course and apply the law to the teaching work of universities, so as to improve the teaching quality and efficiency of teachers.

By utilizing data mining technology to analyze students’ performance in all courses, we can identify correlations and patterns in the performance of different courses and then predict students’ performance in a particular course based on the patterns discovered. In order to apply the results of the prediction to the actual teaching, a teaching system with an integrated module for performance prediction is designed and utilized by teachers in their daily instruction. This system permits the analysis of student academic performance and the forecasting of performance trends [3]. The system can, on the one hand, provide students with targeted advice on their learning process, and on the other hand, it can improve the effectiveness of teachers’ instruction, thereby enhancing the effectiveness of teaching.

In the past, the assessment mode of vocal lessons was almost always fixed, with students singing onstage and the assessment teacher observing and listening onstage, mostly with piano accompaniment, which frequently made students too nervous, making it difficult to sink their breath and causing their voices to become very tight and, in extreme cases, out of tune. Adjusting the accompaniment method of vocal assessment to achieve a more objective assessment effect can be done by allowing students to choose their own accompaniment band, using a small band or self-playing guitar, and so on, to reduce students’ nervousness [4], and allowing students to choose their own performance form, clear singing, solo singing, combination performance, and so on. Regarding the existing pop vocal assessment reform, diversified assessment forms are more popular with students, and the greater the space for free play is, the more the students can fully display their personal characteristics and styles and present daily learning outcomes in a more effective manner. In addition, pop vocal teaching cannot be conducted solely in the piano room, and the dependence of students on the piano room must be gradually weakened so that students do not become accustomed to practicing singing in the piano room in a relaxed and comfortable manner and become nervous when they are not accustomed to the stage environment, which will affect the voice state. This requires teachers to move the teaching classroom to the stage so that students can become accustomed to the stage environment.

This paper’s primary objective is to analyze the detailed learning data of our students over the past few years, identify the factors that have the greatest influence on students in certain specialized courses, and develop a correlation rule; on this basis, a teaching system based on achievement prediction is created. Based on the actual implementation of the C4.5 algorithm to predict student performance in our school, the algorithm was enhanced. The students’ performance in their major courses was influenced by what they learned in their prerequisite courses. The categorization rules will be used to predict students’ performance in vocal singing courses, to alert students to problems in their learning process, and to provide teachers with suggestions for improving the quality and effectiveness of instruction. The categorization rules are incorporated into the design of the system to create a blended learning system based on the prediction of student achievement. Online testing, intelligent scripts, in-class learning, performance prediction, and notification are the primary features of the system.

2. Related Work

2.1. Current Status of Vocal Performance Teaching. First, the fundamental singing skills of students must be improved. Regarding the singing skills of popular vocal music students in colleges and universities at the present time, a large proportion of students have weakened the practice of basic skills with the depth of vocal singing teaching content, which has a direct impact on the singing skills of students. The practice of fundamental skills must accompany the entire professional learning process, which is an essential component of teaching popular vocal music [5]. Second, there is a lack of fluency in pop vocal singing performances. Teenagers are more likely to use pop singing to express their inner thoughts and feelings as a result of the rapid development of the pop vocal singing form, which attracts an increasing number of positive and upbeat musical styles. Numerous contemporary pop vocal students do not sing with their natural voice, resulting in poor vocal performance fluency, which affects the aesthetic perception of pop music works as a whole and does not fully reflect the characteristics of pop music. Thirdly, the ability to process high and low tones must be enhanced. Current performance of many students in the singing process did not master the correct way of high and low tone processing, in addition, the middle and low parts are also an important part of the entire pop vocal singing, and students are often accustomed to singing in the high register to deal with the middle and low parts, but a single reliance on natural tone and pitch is unable to achieve the desired singing effect. In order to achieve a complete expression of the pop music works, it is necessary for the singer to integrate and articulate their bass and soprano singing skills in their singing expression.

Second, the basic training of pop vocal teaching is as follows:

(a) Breathing training: pop vocal singing is based on the combined thoracoabdominal breathing method; this breathing method requires the singer’s mouth and nose at the same time and intercostal muscles with the expansion of the waist around the rib cage to move up, while it requires the diaphragm contraction to move down while expanding the chest cavity around so that breathing can breathe a certain amount of breath into the deep lungs in the shortest possible time. The singer’s breath control during the performance depends on the joint action of the abdominal and lumbar muscles.
(b) Language training bite and the correctness of the way of spitting out words largely affect the singing effect. There are two types of bite, soft bite and hard bite, and the application of flexible penetration can make the emotion penetrate into singing more naturally. The key to bite and spit is pronunciation, and refining pronunciation requires analyzing the head, belly, and tail of each word and finding the connection between them. In order to ensure the consistency of lines and tones when singing, the pronunciation of each word should be standardized to achieve the right amount of front and back, looseness, and tightness.

(c) Vocal exercises: vocal exercises are diversified in ways and means, and there is no fixity. The way of vocal practice should follow the natural development, protect the function movement of vocal organs, and determine the way of vocal practice based on the principle of teaching and differentiation. The control of the progress of vocal exercises should be easy first and then difficult, from shallow to deep, and the strength in a prolonged tone should be consistent. There are many types of vocal exercises, such as skipping, legato, lengthening, and staccato. No matter how easy or difficult they are, teachers need to make appropriate deployment in the corresponding training stage and expand the training on the basis of the original vocalization.

Third, optimize the strategy of pop singing vocal teaching.

2.1.1. Let Students Master the Vocal Principle. High and low tones are important components of pop music works; sublimating music itself means coloring music, but without proper handling of high and low tones, the whole music works will lose color, and it will be difficult for listeners to have emotional resonance when enjoying songs. Most good singers are proficient in mastering and applying high and low notes in their performances. Effective learning of high and low notes is based on understanding the principles of articulation, which is an important prerequisite for proper mastery of specific articulation strategies and techniques. In addition, teachers should communicate with students as equals in vocal music teaching, create a classroom atmosphere for students to perform easily and naturally, and encourage students to study the creation of popular music. Traditional teaching focuses on emotional experience and penetration training for students while ignoring the important foundation of vocal music theory, which will lead students to receive subjective vocal education that will limit the development of their personality in later singing. So in the vocal principle teaching, attention should be paid to the matching of education teaching and students’ pronunciation characteristics, and students themselves should base on their own timbre characteristics and choose their own suitable vocal style according to the vocal principle so as to improve their own emotional expression ability and singing ability.

2.1.2. Improve the Singing Curriculum. Teachers will refer to a large number of textbooks or music resources in their daily courses in order to let students experience a variety of styles of vocal works, and the perfection of curriculum settings largely affects the effect of vocal teaching. The perfection of singing curriculum mainly has the following two points: one, based on a large number of teaching practice summaries, it is known that the professionalism and distinctiveness of textbook selection determine the value of students’ vocal learning materials and the role of teachers’ teaching reference materials, so the selection of professional and unique textbooks is a basic requirement for perfecting the curriculum, which can effectively promote the healthy development of professional vocal education [6]; second, under the influence of the long-term traditional music singing style, the setting of professional vocal music classes needs to match the teaching system with strong characteristics, such as simulated classes, miniclasses, and inverted classes, which have wide application at this stage, which can greatly promote the realization of the professional development goals of colleges and universities.

2.1.3. Improving Students’ Own Cultural Literacy. Music teaching should not only provide students with continuous singing skills training but also improve students’ cultural literacy. The improvement of cultural literacy and musical aesthetic ability is a way to accomplish the goal of comprehensive literacy, which needs to be based on the actual situation and targeted cultural literacy improvement according to the students’ existing knowledge reserves. Although pop music singing has obvious popular characteristics, it is necessary to pay attention to the cultural connotation of the sung works and the singers themselves in addition to the performance form, which requires starting from the performers to improve their cultural literacy, so that they have comprehensive qualities including singing skills, rich emotion, and cultural literacy at the same time [7]. Therefore, the teaching of popular vocal music in colleges and universities should focus on professional music knowledge and singing skills on the one hand and tell students the cultural background and connotation hidden behind the professional knowledge on the other hand, and only a deep and thorough understanding of the culture behind the musical works can make the song singing performance more accurate and promote the comprehensive quality of students. At the present stage of music teaching in colleges and universities, students who have both cultural literacy and singing ability and skills perform better in the interpretation of works and subject assessment, the key of which is that students can accurately grasp the cultural background behind the music works so that they can have more outstanding effects in the expression of emotion and use of skills.

2.1.4. Pay Attention to the Cultivation of Innovative Musicality. Students’ perception of music in daily song singing is a musical sense, and this perception of music is often expressed in the performer’s self-singing style. The
cultivation of innovative music sense needs to start from the following aspects: First, cultivate the sense of pitch. This helps students to understand their own feelings and perceptions of music melodies by positioning themselves in their singing. It requires daily training, independent practice, and perception to gradually cultivate. Secondly, students need to develop the accuracy and stability of their "sense of beat" in order to be able to quickly and accurately identify different beats and to have the ability to remember the original beat with special accent changes. Third, cultivate a sense of rhythm. The cultivation of the sense of rhythm requires long-term training to improve, and the guidance of teachers is indispensable in this process. Fourth, cultivate the sense of force and speed. The performance of the sense of force and speed in singing is the transition, change, and turn of the music content, and the sense of force and speed in the song singing can be controlled well in order to better express the lightness and urgency of the music works and the emotional ups and downs.

2.1.5. Optimize the Assessment Mode of Popular Vocal Music Teaching. In the past, the assessment mode of vocal music class was almost fixed, students sang on the stage, and the assessment teacher observed and listened to them, mostly with piano accompaniment; this simple and stereotypical assessment mode often made students too nervous, which led to difficulty in sinking breath, very tight voice performance, and, in serious cases, out of tune. In order to make the assessment method achieve a more objective assessment effect, one can adjust the accompaniment method of vocal assessment, allowing students to choose their own accompaniment band, using a small band or self-playing guitar and so on, to reduce students' nervousness [8], and allowing students to choose their own performance form, clear singing, solo singing, combination performance, and so on. As far as the existing pop vocal assessment reform is concerned, diversified assessment forms are more popular among students, and the greater the space for free play is, the more the students can fully display their personal characteristics and styles and better present their daily learning outcomes. In addition, pop vocal teaching should not only be carried out in the piano room but also slowly weaken the students' dependence on the piano room so that they will not get used to practicing singing in the piano room relaxed and freely and become nervous when they are not adapted to the stage environment, which will affect the voice state, which requires teachers to move the teaching classroom to the stage, so that students can get familiar with the environment in their daily practice and use the accompaniment to achieve satisfactory teaching results.

2.2. Decision Tree Algorithm. With the development of database technology, the ability to obtain data has been greatly improved, but it is not possible to visually discover the underlying patterns from the data to predict the trend of events. Data mining technology has become one of the hot topics of research because it can help people extract the hidden information and knowledge from a large amount of data and use them for their own benefit. With the development of data mining technology, we can gradually solve tasks including classification, cluster analysis, time series model, association analysis, prediction, and bias analysis. Among them, the classification problem is the most frequently used task in data mining. The most widely used methods for classification problems include decision tree algorithm, K-nearest neighbor algorithm, and support vector machine algorithm.

Among the above-mentioned algorithms, the decision tree algorithm is one of the more widely used classification algorithms because of its simple and practical description. The decision tree algorithm is capable of classifying large-scale data according to certain classification rules and finding valuable information in it. The decision tree originated from the conceptual learning system CLS proposed by Hunt et al., who first proposed the use of decision trees for classification. The most typical decision tree algorithm is the ID3 algorithm based on information gain proposed by Qishan [9], which is the first to link the decision tree with information theory and to use the information gain as the genus selection measure in a pioneering way. However, when using the information gain as the selection criterion for the attributes, the computational characteristics of the information gain lead to a tree structure that favors the attributes with more values, resulting in an unsatisfactory decision tree construction; moreover, the algorithm can only handle data with discrete attribute values, not those with continuous attributes and default values. The top-down nonregression strategy is adopted, and the test property with the highest information gain rate is selected as the classification property.

The application of data mining techniques to the field of teaching and learning has been gradually realized in recent years with the continuous development of the Internet. In one of the more influential papers published in 2007, the German scholar Hans-Peeter Krikgel et al. illustrated the broad prospects of data mining technology in education. The exploration of student-oriented learning motivation and student achievement model based on rough set theory is a very important research [10]. The system uses learning platforms such as Mobile for log analysis. In general, foreign scholars have studied the practice and application of data mining more extensively and have focused on the integration of data mining technology with higher education. Many studies have been conducted on the integration of data mining technology with higher education.

The research on the introduction of data mining technology into education in China is relatively late compared with that in other countries, but there are still many researchers who have made achievements in this area. Weka was used as an experimental platform to analyze students' professional development data by modeling C4.5 algorithms to find patterns in students' behavior hidden in their performance information [11]. Cao et al. used the decision tree algorithm to analyze the factors affecting the performance of exams, and based on the data mining, a set of rules was generated to lay the foundation for the quality of teaching. Xu et al. applied data mining techniques to a computer-level
grading system and used the K-method to analyze students’ test scores, thus realizing a grasp of different students’ learning situations. In general, the use of data mining as a tool to guide education and teaching is conducive to improving the efficiency and effectiveness of education, and therefore the integration of data mining into education and teaching is a necessary trend for the future of education.

3. Key Techniques and Theory

This section introduces the techniques related to system construction and the basics of decision trees.

3.1. Struts2 Framework. With the rapid development of Internet applications in recent years, the user interface logic changes much more frequently than the business logic in Web development. In this case, just changing the display without changing the data and business logic is very difficult to achieve in practice, so almost all Web development should use the MVC design model. As shown in Figure 1, the MVC model achieves separation of model, view, and controller, which greatly improves the maintainability of the code, facilitates code reuse, and makes the physical structure of the Web more reasonable [12].

The model is mainly responsible for the processing of business logic and interaction with the database. It is mainly responsible for encapsulating the data related to the business logic, so it has direct access to the data; the view is only responsible for the display and presentation of the data and has no control over the logic; the controller is the core of the system; it can control the process by capturing and forwarding the requests sent by the user.

The framework is based on the MVC model, which is a free and open-source application framework. The general processing flow is as follows:

1. The client sends a request to the servlet capacitor initialization.
2. The request is filtered by filters, such as Action-ContextCleanUp and SiteMesh.
3. The core controller Filter Dispatcher is invoked and sends an interrogation to the Action Mapper. Based on the result of the interrogation, it is decided whether to invoke Action or not. If an Action needs to be invoked, Filter Dispatcher passes the request to ActionProxy.
4. The ActionProxy uses the Configuration Manager to create an Action instance that uses the Action class to be invoked [13].
5. Call back the execute method of the Action, which operates on the database according to the parameters brought by the request.
6. Information about the result of the processing of the execute method of the Action will be output to the browser.

3.2. Decision Tree Categorization Techniques

3.2.1. Structure of Decision Trees. The origin of the decision tree algorithm is the conceptual learning system, which is the precursor of many subsequent decision tree algorithms, on which many decision tree algorithms are derived with improvements. The decision tree algorithm is an inductive learning algorithm that can deduce the specific representation of the decision tree and the classification rules from a set of data samples that are not regular [14].

The process of constructing a decision tree using the decision tree algorithm is recursive from top to bottom. Specifically, for the classification problem, this process is actually a process of classifying instances based on their characteristics. For example, Figure 2 shows a decision tree in which A, B, and C are the names of the properties, and ai, bi, and c1 are the values of A, B, and C. When the value of the property A is ai and the value of the property B is bi, it belongs to the first class, and the rest of the branches of the tree follow this pattern [15].

The data are preprocessed before constructing the decision tree, where all data are labeled and divided into two parts: one for the training data and the other for the test data [16]. The training data are used to construct the decision tree, while the test data are used to verify the correctness of the decision tree. From Figure 2, we can see that the
constructed decision tree is divided into root nodes, internal nodes, and leaf nodes. The root node and the internal node are the properties of data or the set of properties, and the leaf node is the result of classification. When using the generated decision tree for classification, the root node is first used to compare the property values, then the branch properties below the node are determined, and finally, the leaf node is used to obtain the classification result. In other words, for each decision tree, the root node and the inner node correspond to a test process of a nonclass genus in the data, each branch below the node corresponds to a test result of a different genus, and the leaf node represents the final classification of the test data. Therefore, from the root node of a tree to each leaf node of the tree, each node corresponds to a classification rule, so that the whole decision tree is a set of classification rules.

The process of constructing a decision tree is a recursive process of selecting the most distinguishing features and classifying all the training data based on these features so that the process of data formation results in the best classification subset. The following is a general description of the construction of a decision tree:

1. First, all the training samples on the root node are selected for their properties, and the best trait attribute is selected as the root node.
2. Based on this characteristic property, all the training samples at the node are classified, and the resulting subset of training samples is the best classification at the node.
3. Determine whether the delineated subset of samples is already considered to be the correct classification if it is an arrested node.
4. If there are still some subsets of samples that cannot be classified correctly, then it is necessary to find the best features for these subsets of samples, continue to split the subsets of samples, and create good corresponding nodes.
5. The whole process is a recursive operation, with the final condition that all the data of the training samples are correctly classified or the traits are empty, so that all the samples can correspond to a leaf node.

Attribute selection measure: in the process of constructing a decision tree, the core problem of the whole algorithm is to select the nodes of the decision tree. For different nodes, the selection of condition properties as their test properties to form the optimal decision tree is one of the main focuses of research. There are three main types of genus selection measures commonly used in decision tree classification algorithms, as shown in Table 1.

Information entropy is a measure of the degree of information confusion. If the mixed distribution pattern of information is homogeneous, the information entropy is relatively high; if the pattern of information distribution shows a uniform distribution, the information entropy is low. In the process of constructing the decision tree, if the distribution of classes in the data subset is uniform and mixed, then the entropy of information is high, and if the distribution of classes in the subset is uniform, then the entropy of information is low. The specific definition is as follows [18].

Let $S$ be a set containing $S$ data samples with $m$ different values of class properties, corresponding to $m$ different classes from $C_i$, $i \in \{1, 2, 3, \ldots, m\}$. Assuming that $S$ is the number of $i$ samples in class $C_i$, the amount of information required to classify a given data object is

$$I(s_1, s_2, \ldots, s_m) = \sum_{i=1}^{m} - (p_i \log p_i).$$

Equation (1) defines the concept of entropy, where $P$ is the probability that any data object belongs to class $C_i$, which can be calculated by $s_i/s$. According to the theory of information, all information is coded by bit, and the logarithmic functions in equation (1) are all base 2.

(b) Information gain is the difference between the original information quantity of the data set and the information quantity after classification, and the calculation process of information gain is as follows. Then, the information shelving of the subset classified by the property $A$ is expressed as

$$E(A) = \sum_{i=1}^{m} \frac{s_i}{s} I(s_i, \ldots, s_m).$$

Then, the information gain obtained after the sample set is divided by the current node according to the genus $A$ is

$$\text{Gain}(A) = I(s_1, \ldots, s_m) - E(A).$$

(c) Information gain rate is the normalization of information gain, and the normalization of information gain is based on the concept of split information. The calculation of split information of property $A$ is shown in equation (4):

$$\text{SplitInfo}(A) = \sum_{i=1}^{m} \frac{s_i}{|S|} \log_2 \frac{|S_i|}{|S|}.$$  \hspace{1cm} (4)

Equation (5) is the equation for calculating the rate of increase of attribute $A$:

$$\text{GainRatio}(A) = \frac{\text{Gain}(A)}{\text{SplitInfo}(A)}.$$  \hspace{1cm} (5)

(d) Gini coefficient measures the impurity of the data classification, and the Gini impurity expresses the probability that a randomly selected sample is misclassified in a subset.

$$\text{Gain}(S) = 1 - \sum_{i=1}^{m} p_i^2.$$  \hspace{1cm} (6)
The Gini coefficient is calculated as shown in equation (6), where $A$ is the probability of the $i$-th class of data in $S$. From the equation, it can be seen that the Gini index is larger if the total data set contains more heterogeneous classes.

4. Experimental Results and Analysis

4.1. Data Preparation. The data comes from the teaching workload management platform written by the author, which primarily serves to record teachers' workload in each semester and count the excess credit hours; the platform has accumulated a large amount of data since 2010 [19]. This practice is merely an attempt to learn more useful information from such data. The workload situation of 36 college instructors in the first semester of 2019 was selected at random from the platform, and six feature values were selected to form the original dataset in Table 2 along with the end-of-year teaching evaluation.

Next, the data set in Table 2 was discretized in accordance with the ID3 algorithm's processing rules: (1) Age: it is divided into three classes according to teachers' age groups, with 1 indicating young teachers (less than 30 years old), 2 indicating middle-aged teachers (between 30 and 45 years old), and 3 indicating older teachers (age older than 45). (2) Education: it is divided according to teachers' educational status, where 1 denotes college, 2 denotes bachelor's degree, 3 denotes master's degree, and 4 denotes doctoral degree. 1 indicates assistant professor, 2 indicates lecturer, 3 indicates associate senior, and 4 indicates senior professor. Since the concern is the quantity of course data, the course names are changed to the number of courses, that is, the number of specialized courses taken during the current academic year. (5) Number of students: the number of classes is divided by the number of courses, with greater than 100 indicating more classes, with 1 and 0 indicating fewer classes. (6) Workload: the workload is divided into four categories based on the actual situation: 1 indicates a workload greater than 140 credit hours, 2 indicates a workload greater than 180 credit hours, 3 indicates a workload greater than 240 credit hours, and 4 indicates a workload greater than 300 credit hours. (7) Evaluation of teaching: the original data of evaluation of teaching is percentage data, and we also discrete it into four grades: 0 indicates failure, 1 indicates competence, 2 indicates good, and 3 indicates excellence. Table 3 displays the processed data.

Table 3: Processed data.

<table>
<thead>
<tr>
<th>No.</th>
<th>Age</th>
<th>Education</th>
<th>Title</th>
<th>Course</th>
<th>Number</th>
<th>Workload</th>
<th>Teaching eva.</th>
</tr>
</thead>
<tbody>
<tr>
<td>cz01</td>
<td>Young</td>
<td>Graduate</td>
<td>Associate</td>
<td>Vocal</td>
<td>321</td>
<td>222</td>
<td>Competent</td>
</tr>
<tr>
<td>ay01</td>
<td>Young</td>
<td>Graduate</td>
<td>High</td>
<td>Vocal</td>
<td>220</td>
<td>289</td>
<td>Competent</td>
</tr>
<tr>
<td>f106</td>
<td>Old</td>
<td>Graduate</td>
<td>Associate</td>
<td>Vocal</td>
<td>288</td>
<td>317</td>
<td>Good</td>
</tr>
</tbody>
</table>

4.2. Generating Decision Trees. Using Python as the development language, Python has a clear syntax structure and a rich set of data types such as lists, tuples, dictionaries, collections, and queues, which are simple to manipulate. In addition, many visualization tools are provided in Python to display the results visually. Table 4 shows the change of information gain of the feature values calculated in constructing the decision tree on the training dataset of Table 3, which can reflect the process of constructing the decision tree during the execution of the program.

Table 4: Information gain of feature values.

<table>
<thead>
<tr>
<th>Property selection measure</th>
<th>Algorithm</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information gain</td>
<td>ID 3</td>
<td>The selection of attributes tends to be multivalued</td>
</tr>
<tr>
<td>Information gain rate</td>
<td>C4.5</td>
<td>The gain rate varies significantly with the recursive process</td>
</tr>
<tr>
<td>GINI coefficient</td>
<td>CART [17]</td>
<td>Model is less stable</td>
</tr>
</tbody>
</table>

4.3. Comparison of Methods. As shown in Figures 3 and 4, compared with the course grade analysis system based on the ID3 algorithm and CART algorithm, this system is faster and has the best results.

The loss convergence diagram is shown in Figure 5. It can be seen that the method in this paper converges faster.

4.4. Comparison of Methods

(1) Title is the key factor. Teachers with the senior title have better teaching evaluation, which is in line with the actual situation of the school.

(2) Because of the data, the number of courses is sampled, and most of the teachers teach 2 courses, so this part of the data is a bit distorted, which is a prunable feature; if the training set is changed, it will reflect a different situation.

(3) Workload is the second key factor. From the decision tree, we can find that a high workload is not good for teaching. The workload is reasonable between 180 and 300 hours per semester, and the small workload is not conducive to teaching.

(4) The educational characteristics are also reflected in the decision tree, and the high educational level does not reflect the high evaluation of teaching, which means that this part of teachers still have potential to
Table 3: Modified training data set.

<table>
<thead>
<tr>
<th>No.</th>
<th>Age</th>
<th>Education</th>
<th>Title</th>
<th>Course</th>
<th>Number</th>
<th>Workload</th>
<th>Teaching eva.</th>
</tr>
</thead>
<tbody>
<tr>
<td>cz01</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>321</td>
<td>222</td>
<td>2</td>
</tr>
<tr>
<td>ay01</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>220</td>
<td>289</td>
<td>2</td>
</tr>
<tr>
<td>f106</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>288</td>
<td>317</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 4: Changes in information gain in decision tree construction.

<table>
<thead>
<tr>
<th>No.</th>
<th>Data set</th>
<th>Age</th>
<th>Education</th>
<th>Title</th>
<th>Course</th>
<th>Number</th>
<th>Workload</th>
<th>Teaching eva.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>28.28</td>
<td>0.082</td>
<td>0.65</td>
<td>0</td>
<td>0.298</td>
<td>0.884</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>59.86</td>
<td>0.026</td>
<td>0.8</td>
<td>0.064</td>
<td>0.023</td>
<td>0.765</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>56.553</td>
<td>0.023</td>
<td>0.32</td>
<td>0.054</td>
<td>0.086</td>
<td>0.654</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>69.368</td>
<td>0.228</td>
<td>0.298</td>
<td>0.098</td>
<td>0.068</td>
<td>0.828</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>23.393</td>
<td>0.053</td>
<td>0.039</td>
<td>0.142</td>
<td>0.039</td>
<td>0.896</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>23.8</td>
<td>0.056</td>
<td>0.024</td>
<td>0.289</td>
<td>0.098</td>
<td>0.863</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>24.008</td>
<td>0.082</td>
<td>0.282</td>
<td>0.269</td>
<td>0.123</td>
<td>0.883</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: The results of different methods on the data set.

Figure 4: The results of the data set.
be explored, and how to motivate this part of teachers will be a problem.

(5) The age characteristics are younger teachers and older teachers rated higher.

(6) The class size situation and the number of courses are similar, with more combined classes and some distortion in the data, which are also trimmable features, and more small classes are selected to learn more.

In the above rules, the key factors affecting teachers' teaching quality were found to be title, workload, education, age, or class size in that order. In the management of teachers' workload, teachers' workload can be reasonably allocated according to these key factors; that is, the courses and workload can be reasonably selected according to the characteristics of teachers in order to improve the teaching management.

5. Conclusion

The paper enhances the C4.5 decision tree in accordance with the system's practical application. For various specialized courses, the enhanced C4.5 algorithm is used to generate the decision tree and the corresponding classification rules. In the “Number Signal Processing A” course, for instance, the improved C4.5 decision tree not only was vastly superior to the conventional C4.5 algorithm in terms of operation time but also had an accuracy rate of 80.3%, which was deemed more trustworthy. Data mining technology can be applied to the actual teaching procedure in order to apply the generated classification rules to the system. The paper analyzes the requirements of the teaching system based on the performance prediction of decision trees and identifies the system’s business and functional requirements. On this basis, the system architecture, functional modules, database design, and technology selection are designed and implemented. Although the paper has conducted extensive research and analysis on the prediction and analysis of undergraduates’ performance in specialized courses, it is the first to demonstrate the application of data mining technology in the teaching and learning process, which has some reference value for our education and students’ examinations. However, due to time constraints and the author’s limitations, there are still some flaws in this study. There are numerous algorithms and tools in the field of data mining that have not been evaluated, and it is possible that there are more suitable algorithms for predicting students’ performance that have not been considered. Despite the fact that the final accuracy can meet the prediction requirements, the paper only uses the C4.5 decision tree for improvement and simulation. After the system is implemented, there will be a great deal of information regarding daily registration, self-study, and exams. In the future, the system logs will be analyzed to determine the factors that influence the final examination results of students, such as their learning attitudes and learning habits.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


