

Retraction

Retracted: Decision Tree Algorithm-Based Model and Computer Simulation for Evaluating the Effectiveness of Physical Education in Universities

Complexity

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Z. Zhang, Z. Zhao, and D. Yeom, "Decision Tree Algorithm-Based Model and Computer Simulation for Evaluating the Effectiveness of Physical Education in Universities," *Complexity*, vol. 2020, Article ID 8868793, 11 pages, 2020.

Research Article

Decision Tree Algorithm-Based Model and Computer Simulation for Evaluating the Effectiveness of Physical Education in Universities

Zhifei Zhang,¹ Zijian Zhao ,¹ and Doo-Seung Yeom²

¹School of Physical Education, Zhengzhou University, (School Headquarters), Zhengzhou, Henan 450001, China

²Department of Physical Education, College of Arts & Physical Education, Gangneung-Wonju National University, Gangneung-si, Gangwon-do 25457, Republic of Korea

Correspondence should be addressed to Zijian Zhao; zjzhao@zzu.edu.cn

Received 26 August 2020; Revised 6 October 2020; Accepted 10 October 2020; Published 29 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zhifei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, the forest algorithm and the decision tree algorithm are mainly used to analyze students' physical education information, course exam results, and student learning data and relevant feature attributes from the online teaching platform. We aim to generate decision trees using the decision tree algorithm for the purpose of generating classification rules, based on which we can find factors that are important to students' physical education performance and form data basis for improving teaching quality to help teaching management and teachers improve teaching methods and adjust teaching strategies. We specifically achieved this objective by constructing a model for assessing the effectiveness of student teaching, the steps of which include data collection and preparation, data preprocessing (data cleaning, conversion, integration), model construction (algorithm training), and algorithm optimization, as well as realizing the simulation results of the model. At the same time, the importance of the relevant attributes of the model is analyzed, and some measures are proposed to improve the universities: the standard of physical education teaching and the corresponding strategies for improving teaching methods. The mainstream development environment is chosen to ensure the complete operation of the project system that integrates learning, operation, and evaluation. The sports virtual simulation experimental teaching system realized in this paper has good functionality, stability, and application benefits in operation and use.

1. Introduction

The report "Quality Physical Education: A Guide to Action for Policymakers," published by UNESCO, states that physical education is the only way to combine the development of physical motor skills with the learning and communication of values and is the ideal path for students to acquire core competencies [1]. The study of physical education in universities is the final stage of being able to receive formal physical education and is an important way to refine and improve their knowledge of physical education, and the development of physical education in universities is a prescribed requirement imposed by school education, as well as meeting the needs of the development of sports [2]. The results of the survey on the physical health of

adolescents found that the lung capacity of university students from recent years is a declining trend. Students also have poor vision, neurological failure, and cardiovascular disease, and some of the problems have seriously affected physical and mental health. Compared with university students in developed countries of the world, the fitness and health of university students need to be improved [3]. Existing studies have concentrated on the development of the core competencies of physical education and are directed at transforming and teaching the physical education curriculum, and most of them focus on the theoretical aspects of the basic education field. These studies neglect the factors that influence the development of the person himself as well as other aspects [4]. Therefore, this study is not limited to the disciplinary boundaries but is based on the development of

core human sports literacy as the starting point, in addition to the study of the strategies for the development of core human sports literacy in higher education [5]. Based on the development of higher education, on top of the rich experience in implementing quality education and under the accumulation of practical teaching, the college students' sports core literacy has been promoted [6]. Many factors are influencing the development of college students' physical education core quality. First of all, physical education, as one of the important influencing factors, is the main way for college students to participate in physical exercise, which is closely related to the professional quality of physical education teachers as well as the school's implementation policy, management system, sports environment, and organization of extracurricular sports activities [7]. Therefore, to be able to better cultivate students' sports core qualities, the whole school should be placed in the cultivation system, seek the cultivation implementation way, and optimize the cultivation strategy.

Pai et al. based their PageRank model on the performance statistics and ranking of student-athletes in basketball, hockey, and other sports to assess the individual and team performance of students in a more comprehensive way and to promote their participation in learning [8]. The quantitative performance assessment was used to define the students' game performance, and the qualitative assessment of the students was implemented with a personalized interview form to facilitate the participants' perception and experience of the game performance [9]. At the level of the structural design of the physical education curriculum, Martin et al. in their paper analyzed the new concept of physical education curriculum in comparison with traditional teaching [10]. On this basis, they discuss the importance of classroom assessment as it evolves. The research points out that the New Physical Education (NPE) curriculum perspective, based on national development and youth health issues, has significant implications for traditional curriculum design, implementation, and assessment [11]. Furthermore, research on assessment for physical education and learning has focused on communication and linkages between schools, communities, and families; for example, Kim et al. studied parents' perceptions of school assessment and the need to improve the school's curriculum based on parents' views [12]. Dieter et al. have further analyzed the physical education model (SEM) to assess students' motivation from the perspective of self-determination theory, stating that students need a certain time to determine their behavior in an educational environment. One of the most effective means of dealing with classroom assessment is the design and use of rubrics [13]. Mao et al. state that the use of rubrics to determine students' overall learning performance can greatly enhance students' motivation [14].

As can be seen, there are quite several articles and writings on assessment for physical education learning. Whether it is about the relationship between learning assessment and curriculum development or the specific implementation and application of learning assessment, it is evident that physical education learning assessment has become an important hotspot in today's education sector. It

is necessary to establish the educational concept of "health first" and to provide adequate physical education classes, to help students enjoy fun, strengthen their physiques, enhance their personalities, and temper their wills in physical exercise. These important discourses have greatly elevated the status of college sports and are important guidelines for college personnel training. Apply the software theory to the evaluation system functional requirement analysis, design the main functional modules, realize the decision support system based on data mining and simulation, improve the work efficiency of the evaluation department, and provide decision guidance through certain mining analysis. The research proposes a simple and effective data warehouse design and implementation method and uses online analysis and processing technology to analyze the data in the data warehouse and provide basic data for subsequent data mining. After studying and researching the commonly used data mining algorithms and understanding the characteristics of different algorithms, a decision tree algorithm suitable for this data mining is selected based on the data studied in this paper, an improved algorithm is used to establish the mining model, and the mining results are analyzed to extract valuable decision information. According to the problems found by applying the decision tree algorithm, the algorithm is studied in depth, analyzed, and compared, combined with the characteristics of the data of this project, and the analysis of data mining and data warehouse and online analysis and processing techniques are used to obtain useful information to provide a meaningful reference for decision making in the evaluation of physical education.

2. Design of a Decision Tree-Based Model for Assessing Teaching Effectiveness in Universities

2.1. Improved Decision Tree Algorithm and Computer Simulation Design. As one of the most commonly used methods for data mining, the decision tree algorithm has been widely used in different fields since its introduction. It has undergone a long process of going from shallow to deep and from simple to complex. The decision tree is a top-down, recursive division that uses a top-down, divide-and-conquer approach, and its basic algorithm is essentially greedy [15]. Starting from the root node, each nonleaf node is found to find an attribute in its corresponding sample set to test the sample set, and the training sample set is divided into several subsamples according to the different results of the test. Each subsample set constitutes a new leaf node, and the above process is repeated for the new leaf node so that the loop continues to reach a specific termination condition [16]. The flexible approach to management adopted by teachers and teaching, which is determined by the characteristics of university work, is to go to class when there are classes and then to have free time when there are no classes. While this is conducive to research and learning, it is also likely to keep certain aspects of teaching, such as lesson planning, homework revision, and lab instruction, out of control. The

decision tree adopts a top-down recursive approach to compare and evaluate the attribute values of nodes within the decision tree and determine the branch down from the node based on the different attribute values. One of the biggest advantages of the decision tree-based learning algorithm is that it does not require the user to acquire a lot of basic knowledge during the learning process, as shown in Figure 1.

The process of constructing a decision tree is divided into two steps: tree building and pruning. The first step is the tree building stage, which selects part of the training data and builds a decision tree by the breadth-first recursive algorithm until each leaf node belongs to the same class. The second step is the pruning stage, which uses the remaining data to check the generated decision tree and correct the errors, and it finally prunes the decision tree and adds nodes until a correct decision tree is built. The decision tree building algorithm is a recursive process that ultimately results in a decision tree, and pruning reduces the impact of noisy data on classification accuracy. In general, the greater the information gain, the greater the “purity improvement” obtained by using features to partition the dataset. Therefore, information gain can be used to select attributes for decision tree partitioning, which is to select the attribute with the greatest information gain.

Information gain is the amount of change in the desired information, which mainly reflects the method of sample feature importance, and there is a positive correlation between the two: the importance of the sample feature increases and decreases with the amount of information gain. Information entropy is mainly used to measure the amount of information, so we define the entropy w of data [17].

$$\begin{aligned} Q_L(w) &= \frac{1}{2} \sum_{i,j} P_{ij} (w^T x_i - w^T x_j)^2 \\ &= \frac{1}{2} \sum_{i,j} P_{ij} (w^T x_i - w^T x_j)(w^T x_i - w^T x_j). \end{aligned} \quad (1)$$

The information gain rate and destination information gain will tend to select the attribute with the larger information gain value. It applies a parameter value to the specification of the information gain, which is calculated as follows [18].

$${}^G D_t^v f(t) = \lim_{h \rightarrow 0} \frac{1}{h^v} \sum_{m=0}^{[t-ah/h]} (-1)^m \frac{\Gamma(v+1)}{m! \Gamma(v-m+1)} f(t-mh), \quad (2)$$

and based on this, the formula for calculating the gain rate can be obtained as follows:

$$m_i(g) = \frac{\text{fit}_i - \text{worst}(g)}{\text{best}(g) - \text{worst}(g)}. \quad (3)$$

There are many algorithms for generating decision trees in data mining, and several typical decision tree generation algorithms are highlighted below. The ID3 algorithm is the most influential and typical in decision tree mining, which

selects test attributes for each nonleaf node in the tree through a set of rules based on information theory, using entropy as the basis for classification, and finally classifies the data into the form of a decision tree [18]. The basic idea of the ID3 algorithm is to use information entropy as a measure of attribute selection of decision tree nodes; each time it first selects the attribute with the most information, i.e., the attribute that can change the entropy value to the smallest, to construct a decision tree with the fastest decrease in entropy, so that the entropy value at the leaf node is 0 [19]. Currently, the set of instances corresponding to each leaf node belongs to the same class. Since the attribute with the highest gain of information is always selected to divide the rules, the algorithm classifies data quickly, the depth of the tree is average, and the division rules are simple. The ID3 algorithm is important in the process of constructing a decision tree: attribute determination and set partitioning. The process of attribute determination is to select the attribute with the largest value of gain information from the selected data as the root node and construct a branch with the attribute value of this node to divide the selected data into several unrelated subsets; after branching the nonleaf node, it will determine the attribute of the subset value twice and then continue to branch until it branches to the leaf node.

C4.5 is a classification decision tree algorithm in machine learning, which is based on the ID3 algorithm and improves on the advantages of ID3 again. This algorithm uses the concept of information gain or entropy reduction to select the optimal division, to better realize the construction of the decision tree: (1) It selects attribute columns based on the information gain rate, overcoming the deficiency of biased selection of attributes with more values when selecting attributes with information gain. (2) It constructs a decision tree. The tree pruning process is completed in the tree construction process. (3) The data of continuous attributes are processed by applying discretization. (4) The tree pruning construction is also used for incomplete data. The ID3 algorithm can only solve the discrete data attributes, but the C4.5 algorithm can handle the discretized continuous attributes well, with the following procedure.

$$Q(w) = \frac{Q_N(w)}{Q_L(w)} = \frac{\text{tr}(w^T X_L N X^T w)}{\text{tr}(w^T X_L X^T w)}. \quad (4)$$

Computer simulation (data visualization), as the name implies, is a simulated representation of data that uses computer graphics and image processing techniques to interpret data through analysis, transformation, and graphic patterns (including animation) in either planar or three-dimensional form, while providing methods, theories, and techniques for interaction [20]. It helps people to view data and the relationships between data more visually. The atlas of computer simulation is composed of a large amount of data. Each data item is the basic unit of the atlas, and the value of each attribute of the data is represented by the multidimensional terms so that the user can observe and analyze the data from different angles to better serve the user and help the user to make decisions. Simulation technology is the data mining results of abstract information, with a

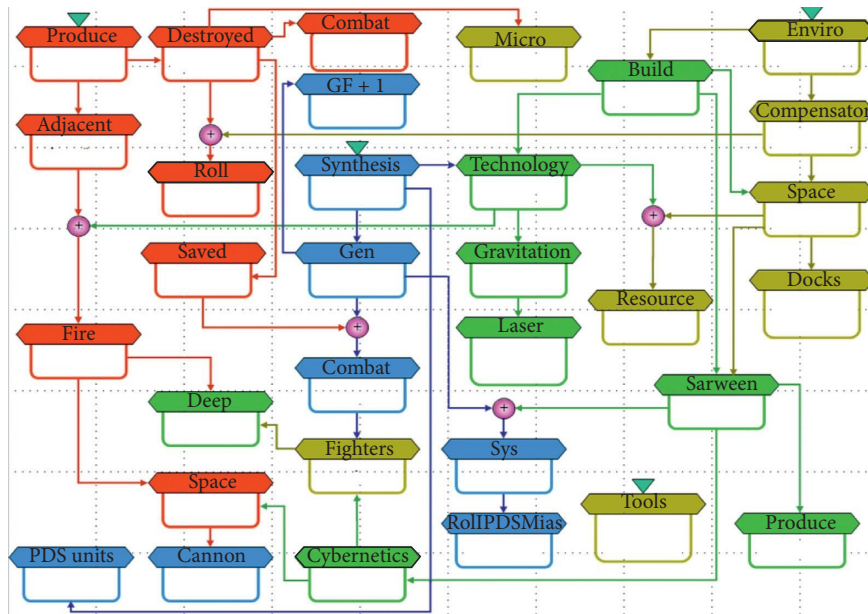


FIGURE 1: Model of improved decision tree algorithm.

simple and intuitive form of expression, to deepen the user's understanding of the meaning of the data, while understanding the interrelationship between the data and development trends. The characteristics of computer simulation technology are as follows. There is strong interaction with the user. The user is no longer simply a receiver of information but can also manage, process, and develop the data. Computer simulation technology can classify, arrange, and present data from multiple perspectives. For example, users can view data time, percentage, rank, and other dimensions. Users can view data and analyze it using simulation charts, histograms, line graphs, puzzles, etc. [21].

Scientific computational simulation refers to the use of computer graphics and image processing techniques to display engineering measurement data, data generated by scientific calculations, and calculation results on the screen and transform them into images. They can also be processed interactively using theoretical methods and techniques. The classification of scientific computing data is relatively extensive and can be divided into structured, unstructured, and mixed data according to different data structures or scalar, vector, and tensor according to data types. Scientific computing simulation technology faces two difficulties in the development process: first, the way to classify the object data by research; second, how to display the simulation object on the screen practically and effectively, making users view it interactively.

2.2. Evaluation Design of the Effect of Physical Education in Universities. Model building is at the heart of data mining, which means determining which algorithm to choose to build the decision tree. The implementation of the algorithm requires certain tools for programming. In this paper, MATLAB tools are used to build the CART decision tree. There are three main reasons for this. Firstly, the research

content of data mining is the learning and application of models and algorithms, and MATLAB is particularly suitable for algorithm development. Because MATLAB can directly call a rich library of mathematical functions to quickly implement algorithms according to the flow steps of the algorithm, the programming workload using MATLAB will be greatly reduced when the computational requirements are the same. The syntax of MATLAB is more intuitive than Python or R, and the learning and programming of algorithms are easier. Secondly, MATLAB has efficient and rich scientific computation functions, including calculus, matrix computation, and symbolic computation, and system simulation applications are very extensive. Thirdly, MATLAB itself is a program development tool with a friendly GUI development function [22]. MATLAB was used to learn the principle of algorithms in depth, and then hands-on programming computer programs were deployed to solve many machine learning problems in practical applications. Using MATLAB is more focused, and the toolbox is mostly developed in C. The tools are all interconnected, and it is easier to get started if you have or understand this data mining knowledge and then apply these tools, as shown in Figure 2.

From the above rules, by analyzing the meaning of the rules in detail, some teaching strategies can be adjusted for the problems that arise, such as the need to have relatively effective teaching methods in the online teaching platform, whether to make it mandatory or set open assignments (conscientious students do open assignments, while lazy students do mandatory assignments) and guide students to learn independently. Other aspects, such as out-of-province students with a poor foundation and students with a good foundation who also fail exams, can be predicted pre-intervention can be made; students with good English scores generally have a positive attitude to learning and are encouraged to study professionally. Since the data is not

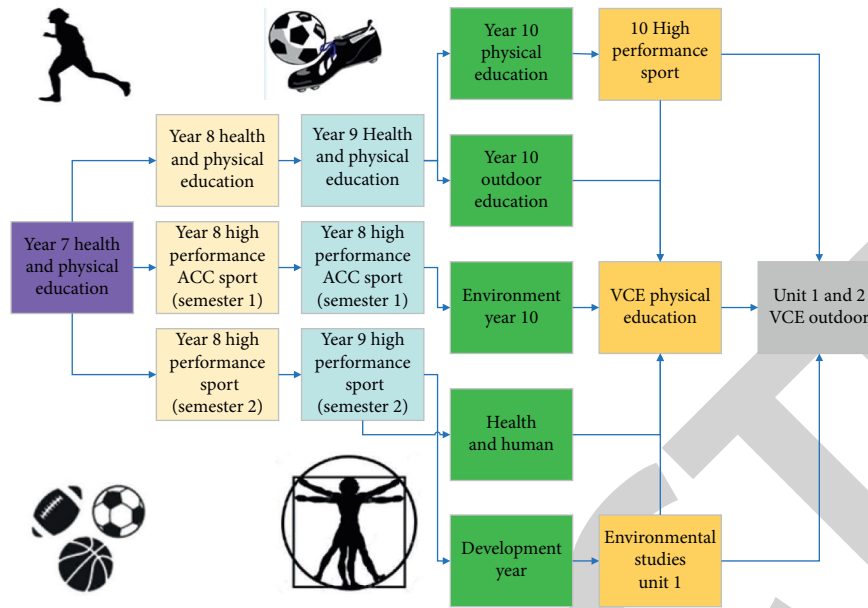


FIGURE 2: The design of physical education effectiveness evaluation.

complete and comprehensive, the results obtained can only serve the major students' local construction advice, but to some extent to make the assessment of student learning effectiveness possible. With the continuous deepening of the new curriculum, curriculum reform is steadily advancing in depth; in the gradual integration of the new curriculum concept and classroom teaching, we have made some achievements and accumulated some experience in the classroom. At the same time, many problems often arise in the classroom, which we did not anticipate in advance, when we must think of a good strategy and find a good way to solve them.

With the development of colleges and universities and the advancement of network technology, the introduction of data mining technology, visualization technology, etc. into the assessment information management system of colleges and universities will provide great convenience and work efficiency. Besides, through data mining technology to analyze student information over the years, you can find some valuable laws and factors that affect the assessment of student reporting, hence more scientific guidance to adjust our plans and decisions. The design of the database system directly affects the quality and operation of the assessment management system. The relative independence, data integrity, and consistency of the database were followed in the design of the database. According to the design features of this system, the traditional database processing mechanism differs greatly from the decision analysis needs and cannot handle online analysis processing, so a data warehouse was established based on the database for the data between different databases. The evaluation management information of this system uses SQL Server 2012 as the software for data storage and management, and this paper only takes the design of the evaluation database as an example, aiming to illustrate the relationship between data tables, databases, and data mining functions and visualization functions.

2.3. Performance Evaluation Design. When performing data mining algorithms, to improve the college's reporting success rate, the assessment dataset is used as the object of study; therefore, the decision tree algorithm is used to analyze the key attributes of the assessment data source that may affect the reporting rate and to identify the factors that are most likely to affect the student reporting, so that these potentially relevant influences can be scientifically applied to provide key decisions for future assessment work. The advantages of decision tree algorithms include easy-to-understand analysis, high classification accuracy, and high execution efficiency, making decision tree algorithms well-suited for use in mining large amounts of data. The commonly used decision tree algorithms are ID3 and C4.5, but the two algorithms are very different in the selection of splitting rules. ID3 selects the classification criterion of information gain, which is to select more attribute values as splitting attributes, and it can only mine nonlinear data. C4.5 evaluation classification criterion is the information gain rate, that is, according to information; therefore, C4.5 can effectively circumvent the problems of ID3. Moreover, through the comparative analysis in Section 2.1, the decision tree algorithm of C4.5 is preferred to be used for data mining for better attribute factors, as shown in Figure 3.

The improved algorithm merges branches with higher and lower entropy values, i.e., branches where the division is not important, effectively reducing the fragmentation problem, improving branch efficiency, and limiting the impact of the overfitting problem. In contrast, this paper studies the promotion of assessment in universities, with diverse student sources, but the number of people assessed in each major is relatively small due to local and other reasons, so this requires a more balanced analysis of the diverse assessment channels and student sources, to prevent analysis bias. The improved C4.5 decision tree algorithm makes better use of memory space, balances the selection of

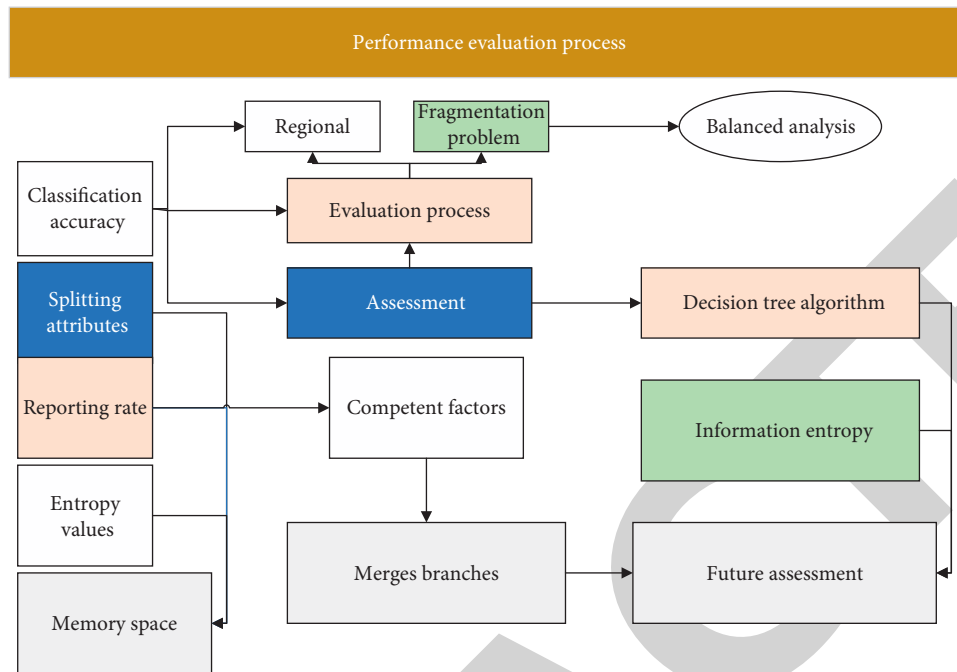


FIGURE 3: Performance evaluation process design.

information entropy, and avoids low or high information entropy caused by the human influence of competent factors, which is more conducive to a more accurate decision tree.

Due to the small data amount selected in the selection of the training set, there may be some deviations in the obtained mining model, and then the training set model is verified through the test set to verify whether the model is accurate, so that the mining model can be further improved and revised, which will be analyzed in the decision tree analysis function module of the system module analysis.

3. Results and Analysis

3.1. Analysis of Assessment Results. It is found that the more the decision tree base classifiers are, the smaller the generalization error will be. The accuracy of the algorithm is close to 98.25% compared to 82.58% for a single base classifier. This is comparable to the accuracy of the CART algorithm. This shows that the decision tree algorithm can predict new data better than the CART algorithm.

In Figure 4, it is found that the prediction error rate of the classifiers is decreasing; i.e., the prediction accuracy rate is increasing. It is more clearly known that the algorithmic accuracy of the decision tree keeps improving with the increase in the number of classifiers. Evaluating the decision tree classification performance cannot be done by simply counting the accuracy of a single breakpoint, but it requires analysis of the classifier's ROC curve, AUC value, and other metrics. The decision tree algorithm is used to validate predictions for 30.25% of the test sample and compare them with the true category. The ROC curve and AUC value are calculated and Figure 5 is obtained. It has an AUC value of

0.7444, although the ROC and AUC metrics are improved by a point over the CART algorithm and the decision tree is more accurate. However, whether the comprehensive performance of this algorithm can still be improved remains to be explored further by the experiment. The ROC (Receiver Operating Characteristic) curve and AUC (Area under the Curve) values are often used to evaluate a binary classifier (the advantages and disadvantages of ROC and AUC (classifier) are discussed here). The ROC and AUC are used in the evaluation of model prediction results when doing medical image computer-aided pulmonary nodule detection in the past. The characteristics of ROC and AUC are briefly introduced here, and how to make ROC plots and calculate AUC values is discussed in more detail.

An analysis of the unsatisfactory experimental results, in terms of the distribution of the dataset, reveals possible causes. One is the problem of data imbalance in the dataset. The target variable classification ratio in the two-category dataset was 586:110 (pass or fail). Afterward, we adjusted the target classification ratio to 407:258 (mean division) for training, made predictions on 30.25% of the test data, and then conducted the ROC curve and AUC value statistics. The expected experimental effect was achieved. The decision tree algorithm can predict the classification more accurately. However, the importance of the attributes of the independent variables cannot be represented as clearly as the decision tree. According to the most important attribute summarized in the decision tree (professional grades), after removing this attribute and retraining the classifier, the false and true rates of the test set can be calculated to obtain the ROC curve Figure 6, which has an AUC value of 0.611 and a decrease of 23.34%, which shows the importance of this attribute. Similarly, the ROC curve and AUC value can be

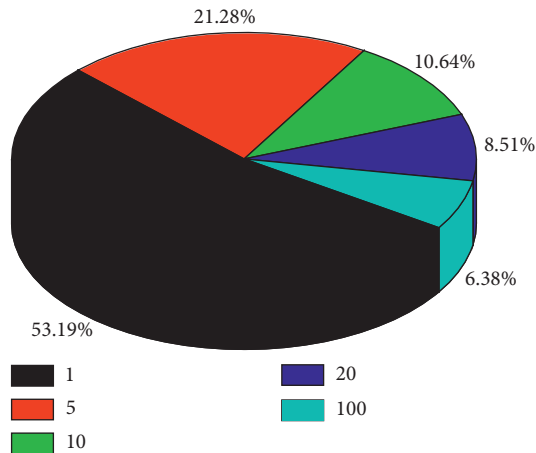


FIGURE 4: Number of base classifiers and generalization error.

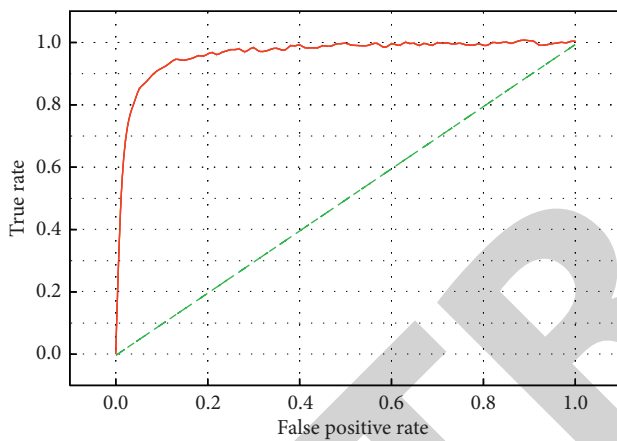


FIGURE 5: ROC of a decision tree on 30% test set.

statistically determined by removing the other attributes. In this way, we can indirectly analyze the characteristics of the most important attributes that affect the target variable.

The AUC of the highest decision tree algorithm is only 0.8441, which is far from excellent performance. The highest decision tree algorithm has an AUC of only 0.8441, which is far from excellent performance. However, the actual classification application in this dataset can predict the classification more accurately and meet the practical application requirements.

The CART and RF algorithms are used to train the data. Firstly, the CART model is used as a single classifier, and the CART classifier is used as a base classifier. The data is trained using the random forest algorithm with different numbers of base classifiers, resulting in four integrated classifiers and a total of five classifiers. The new data table “data” is predicted using each classification model, and the predicted results are compared with the actual classification labels, as shown in Figure 7.

The results show that for both decision tree (CART) and random forest (RF) algorithms, by constructing a random forest algorithm based on CART classifier, the random forest

algorithm integrated with a different number of base classifiers enhances the classification evaluation indexes, in which the true rate, precision-P, recall-R, and F1 value are improved, and the classification performance of the model reaches the expected results. The accuracy of the random forest algorithm is 0.981 for the 5-tree classification model, 0.99 for the 10-tree classification model, and 0.991 for the 20-tree and 100-tree classification models. The classification accuracy has reached the ideal state. As for the comprehensive performance ROC and AUC metrics of the classifier, a comparison has been given in the previous section and will not be repeated here.

In the process of predicting student achievement using classification algorithms, the performance of the final classification model is related to not only the criteria of algorithm goodness and split attribute selection, but also the collection of raw dataset and data preprocessing. The key to classifying and predicting student performance is the need to filter out the main attributes that influence student course performance from the attribute characteristics of the raw data, use this influencing factor as a candidate attribute set, and construct a classification prediction model with student test scores as the target variable. It is found that the attribute set of the raw data may not be very comprehensive and may lack some of the main feature attributes; then, the accuracy of the constructed classification model will not be very high. Therefore, the algorithm is not a panacea; there is no perfect algorithm in the world, and perfect data needs to be collected to implement it. Anomalous data in the dataset can also affect the classification accuracy of the model. It affects a single decision tree model more, while integrated combinatorial models are relatively less easy to affect, and it has a strong generalization ability. However, if multiple trees are trained on the same data, it is also easy to get strongly correlated trees, and then the effect is rather bad, which is all related to the choice of the sample set and attribute selection.

3.2. Analysis of Computer Simulation Results. According to the system test case design scheme, the virtual simulation experimental system of “Human Movement Ability Assessment and Fitness Path Design” was tested. 48 test cases were set up and the test report is shown in Table 1.

Among them, 18 user interface test cases, 20 functional test cases, and 10 performance test cases passed the test at a rate of 100%, which shows that the system has passed the application test and can be put into teaching application, as shown in Figure 8.

The interaction of the system is the basic characteristic of virtual simulation experiments. According to the constructivist theory, keeping good interaction in the learning process can effectively stimulate learners’ motivation and enhance the learning effect. Ease of use is an important dimension to evaluate the logic and science of the virtual simulation experimental teaching system in terms of functional design, which can provide important instructions for the design and optimization of the system. Besides, the vivid experimental scenarios and easy-to-use operation

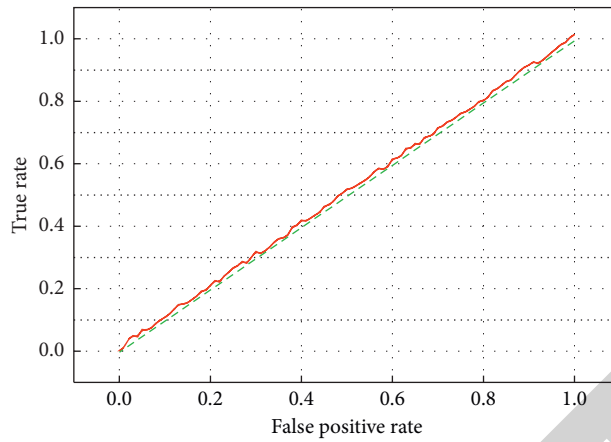


FIGURE 6: ROC curve with the main independent variables removed.

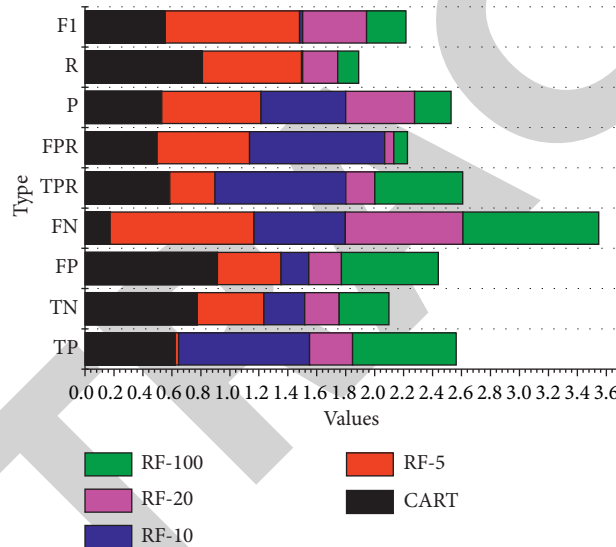


FIGURE 7: Performance comparison of CART and RF classifiers.

TABLE 1: Test cases.

Test cases	Number of questionnaires issued	Number of questionnaires withdrawn	Number of valid questionnaires	Recovery rate (%)	Efficiency (%)
Numbers	12	16	20	93.65	94.86
Parents	14	14	20	87.47	86.74
Total	14	15	19	88.47	85.78

procedures are also beneficial to the efficient operation of the knowledge learning process, which is an effective extension of the traditional experimental sports education.

The usefulness of the system is one of the indicators to evaluate the teaching effect of sports virtual simulation experiments. The goal of sports virtual simulation experiments is to meet the actual needs of sports experimental teaching and improve the teaching effect of the experiments. By evaluating the usefulness of the system, we can clearly understand the learners' knowledge of virtual

simulation experiments, to provide a basis for the assessment of the practical effect and teaching significance of sports virtual simulation experimental teaching system. The motivation of the system is the psychological feedback of the learners after using the virtual simulation experimental system. The high level of motivation indicates that the learners recognize the value of the virtual simulation experiments in terms of learning effect; on the other hand, it also indicates that the system can dig deep into the learners' learning needs to strengthen their desire for

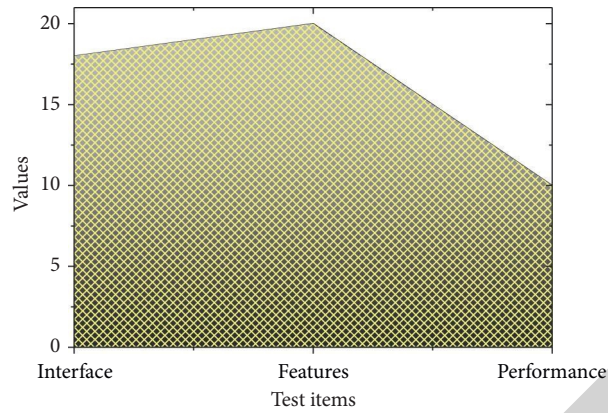


FIGURE 8: Test report of virtual simulation experimental system.

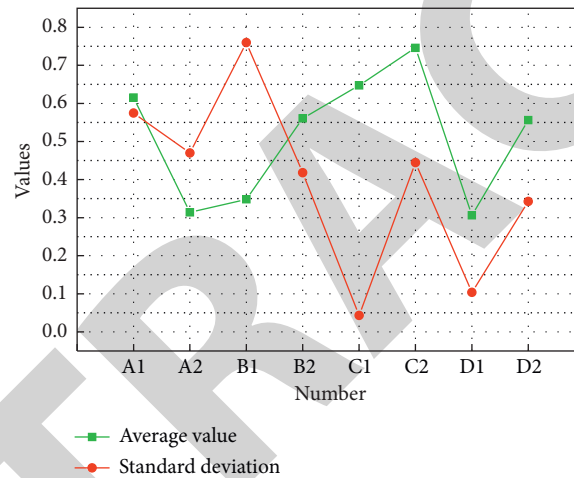


FIGURE 9: Descriptive statistics of sports virtual simulation experiment.

learning sports knowledge, which reflects the guiding role of the virtual simulation experiments in sports to the learners. The development of virtual simulation experimental teaching in higher education is an important content of higher education informatization. This paper points out the necessity of virtual simulation experimental teaching according to the experimental practice of higher education in China, analyzes the problems of the current virtual simulation experimental teaching project, and proposes the sustainable development mechanism of virtual simulation experimental teaching project.

The application function and teaching effect of sports virtual simulation experimental teaching are evaluated by a Likert five-point scale and the specific results of statistical analysis are shown in Figure 9. The above results show that the test subjects are highly satisfied with the virtual experiment of “Human Movement Ability Assessment and Fitness Path Design,” with an average score of more than 4, which indicates that the four dimensions of interactivity, usability, usability,

and motivation of the system are at a relatively good level, and the system is recognized and praised by the test subjects.

From the evaluation feedback of interactivity, ease of use, usefulness, and motivation, we can see that the system can build a virtual simulation teaching scenario with good interactivity, which enables the test subjects to use the virtual experimental equipment and instruments easily, freely, and fluently to learn and explore sports knowledge independently and improve their sports knowledge, practical application ability, and learning motivation.

In this paper, two decision tree algorithms, ID3 and C4.5, are used to select the same training dataset to construct models of student reporting situations, and then the same test dataset is used to evaluate the performance of these two different models. The best performing algorithm is shown in Figure 10.

The accuracy of the two algorithms was calculated according to the formula for the accuracy assessment index, and Figure 10 shows that the accuracy of C4.5 is significantly higher than that of the ID3 algorithm.

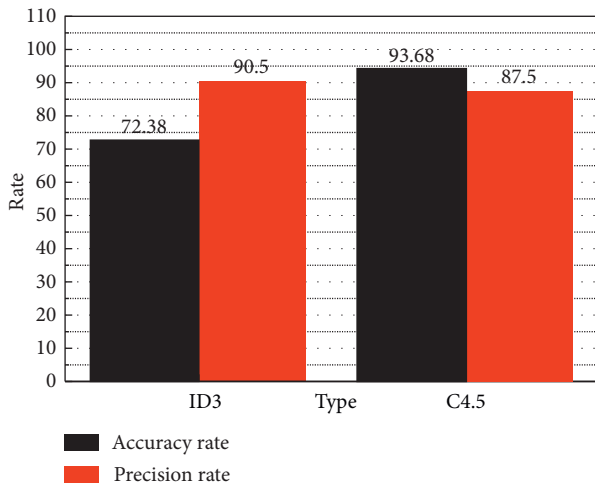


FIGURE 10: Accuracy rate and precision rate.

4. Conclusion

In this paper, the CART algorithm is applied to the analysis and prediction of students' basic information, course test scores, and e-learning platform learning behavior characteristics. In addition, a CART model is constructed based on the improved decision tree algorithm to study the effectiveness of physical education in universities and to study its evaluation model and computer simulation. The CART identifies rules for correlating student learning factors, and some pedagogical strategies can be adjusted, such as the need to have a relatively effective teaching method in the e-learning platform, whether to make it mandatory or set open-ended assignments to guide students' independent learning. In addition to guiding the importance of foreseeing the building of the profession in terms of student and professional teaching, some degree of assessment of the effectiveness of student learning may be achieved. There are many different methods of data mining, among which decision trees and random forests have a very wide range of applications and practical implications. This is because they are theoretically clear, easy to understand, relatively computationally modest, and highly accurate. In this paper, the application of CART algorithms to the analysis and prediction of the assessment of student effectiveness in physical education is significant.

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 known financial conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] N. Segev, M. Harel, S. Mannor, K. Crammer, and R. El-Yaniv, "Learn on source, refine on target: a model transfer learning framework with random forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1811–1824, 2017.
- [2] P. A. Vagefi, D. Bertsimas, R. Hirose, and N. Trichakis, "The rise and fall of the model for end-stage liver disease score and the need for an optimized machine learning approach for liver allocation," *Current Opinion in Organ Transplantation*, vol. 25, no. 2, pp. 122–125, 2020.
- [3] M. R. Nachiappan, V. Sugumaran, and M. Elangovan, "Performance of logistic model tree classifier using statistical features for fault diagnosis of single point cutting tool," *Indian Journal of Science and Technology*, vol. 9, no. 47, pp. 1–8, 2016.
- [4] S. M. M. Rubiano and J. A. D. Garcia, "Analysis of data mining techniques for constructing a predictive model for academic performance," *IEEE Latin America Transactions*, vol. 14, no. 6, pp. 2783–2788, 2016.
- [5] X. Ke, L. Qi, and C. Zeng, "A partitioned and asynchronous cellular automata model for urban growth simulation," *International Journal of Geographical Information Science*, vol. 30, no. 4, pp. 637–659, 2016.
- [6] S. Figini, R. Savona, and M. Vezzoli, "Corporate default prediction model averaging: a normative linear pooling approach," *Intelligent Systems in Accounting, Finance and Management*, vol. 23, no. 1-2, pp. 6–20, 2016.
- [7] S.-W. Lin, Y.-T. Liu, and M. A. Jerusalem, "Bayesian reliability analysis of a products of probabilities model for parallel systems with dependent components," *International Journal of Production Research*, vol. 56, no. 4, pp. 1521–1532, 2018.
- [8] P.-F. Pai, L.-C. Chen, and K.-P. Lin, "A hybrid data mining model in analyzing corporate social responsibility," *Neural Computing and Applications*, vol. 27, no. 3, pp. 749–760, 2016.
- [9] S. A. Mokeddem, "A fuzzy classification model for myocardial infarction risk assessment," *Applied Intelligence*, vol. 48, no. 5, pp. 1233–1250, 2017.
- [10] J.-P. Martin and D. Germain, "Dendrogeomorphic reconstruction of snow avalanche regime and triggering weather conditions," *Progress in Physical Geography: Earth and Environment*, vol. 40, no. 4, pp. 527–548, 2016.
- [11] G.-T. Gwak, S.-H. Ahn, J.-h. Kim, Y.-s. Weon, and O.-Y. Kwon, "Prediction model for the risk of scapular winging in young women based on the decision tree," *Physical Therapy Korea*, vol. 27, no. 2, pp. 140–148, 2020.
- [12] H.-K. Kim, K.-H. Choi, S.-W. Lim, and H.-S. Rhee, "Development of prediction model for prevalence of metabolic syndrome using data mining: Korea national health and nutrition examination study," *Journal of Digital Convergence*, vol. 14, no. 2, pp. 325–332, 2016.
- [13] W. B. Dieter, J. P. Collins, and A. A. Guccione, "Predicting outcomes within an innovative post-acute rehabilitation model for older adults," *BMC Geriatrics*, vol. 19, no. 1, pp. 1–9, 2019.
- [14] Y. Mao, M. Zhang, P. Sun, and G. Wang, "Landslide susceptibility assessment using uncertain decision tree model in loess areas," *Environmental Earth Sciences*, vol. 76, no. 22, pp. 1–15, 2017.
- [15] G. Tao, H. Song, J. Liu, J. Zou, and Y. Chen, "A traffic accident morphology diagnostic model based on a rough set decision tree," *Transportation Planning and Technology*, vol. 39, no. 8, pp. 751–758, 2016.

- [16] M. S. Yorgun and R. B. Rood, "A decision tree algorithm for investigation of model biases related to dynamical cores and physical parameterizations," *Journal of Advances in Modeling Earth Systems*, vol. 8, no. 4, pp. 1769–1785, 2016.
- [17] Y. Tanoue and S. Yamashita, "Loss given default estimation: a two-stage model with classification tree-based boosting and support vector logistic regression," *Journal of Risk*, vol. 21, no. 4, pp. 19–37, 2019.
- [18] X. Wang, Z. Guo, W. Zhang et al., "Study on the decision tree model for risk factors of vascular cognitive impairment," *Chinese Journal of Behavioral Medicine and Brain Science*, vol. 26, no. 6, pp. 534–538, 2017.
- [19] J. Zhao, "Study on the application of classification tree model in building the risk model for nosocomial infection in critically ill patients of emergency department," *Journal of Postgraduate Medicine*, vol. 40, no. 9, pp. 791–794, 2017.
- [20] D. K. Nagthane and A. M. Rajurkar, "An improved diagnosis technique for breast cancer using LCFS and TreeHiCARE classifier model," *Sensor Review*, vol. 39, no. 1, pp. 107–120, 2019.
- [21] R. Prasetya, "Data mining application on weather prediction using classification tree, naïve bayes and K -nearest neighbor algorithm with model testing of supervised learning probabilistic brier score, confusion matrix and ROC," *Journal of Applied Communication and Information Technologies*, vol. 4, no. 2, pp. 25–33, 2020.
- [22] S. Jia and Y. Pang, "Teaching quality evaluation and scheme prediction model based on improved decision tree algorithm," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 13, no. 10, pp. 146–157, 2018.