

## Research Article

# Analysis and Evaluation of Sports Effect Based on Random Forest Algorithm under Big Data

Kai Liang <sup>1</sup> and Dongdong Zang<sup>2</sup>

<sup>1</sup>Department of Physical Education, Jiangsu Vocational College of Agriculture and Forestry, Zhenjiang, Jiangsu 212400, China

<sup>2</sup>Economics and Humanities College, Jiangsu Vocational College of Agriculture and Forestry, Zhenjiang, Jiangsu 212400, China

Correspondence should be addressed to Kai Liang; [liangkai@jsafc.edu.cn](mailto:liangkai@jsafc.edu.cn)

Received 11 May 2022; Revised 29 June 2022; Accepted 21 July 2022; Published 29 August 2022

Academic Editor: Jiafu Su

Copyright © 2022 Kai Liang and Dongdong Zang. 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.

Under the background of big data, all walks of life have carried out in-depth informatization construction. As an important part of national education, the informatization construction of universities cannot be ignored. In recent years, the state has promulgated various policies and regulations, which provide a guarantee for the normal development of school physical education and can improve the current situation of college students' declining physical health to a certain extent. This study attempts to set some specific indicators to promote the better implementation of college sports so that it can supervise and inspect them from all aspects during the actual development of college sports and provide a standard for measuring the implementation effect of college sports. Based on the RF (Random Forest) algorithm, this paper puts forward an evaluation algorithm of students' sports achievements, which can be used to predict students' sports achievements and at the same time, find out the factors that affect students' learning and rank them in importance. The results show that the confidence level of sports effect evaluation by this method is high, and the average confidence level is above 0.96. *Conclusion.* This method has improved the effect of sports effect evaluation, thus effectively guiding sports skills training and improving sports skills.

## 1. Introduction

Vigorously developing college students' physical education is a major event to promote the construction of a healthy China and realize the Chinese dream. It is necessary to make college students love sports and participate in sports as a living habit so that college students can have a healthy physique, good interpersonal relationships, and strong character. It is necessary to mobilize all social forces to protect the health of college students. The state has made constant regulations on the quality and quantity of sunshine sports, which is enough to show that the state attaches importance to college students' participation in sunshine sports and cares about students' physical health. The research shows that there are many declining trends of college students' physical health in China, and one of the most important reasons is that they do not cultivate good sports habits and cannot do extracurricular exercises by scientific methods or means.

Sports achievement is mainly a measure of whether there is a gap between students' physical fitness and expected goals and the means to measure this gap. It can only show the results of students' physical exercise but cannot reflect the causes of such results. Yang and Li introduced the classic Apriori algorithm and the famous DT (Decision Tree) ID3 algorithm of association rules [1] and used Apriori association rules algorithm to mine the influence degree of a course's excellence on other courses. Attar et al. put forward a design idea of multistrategies, combining DM technology with statistical analysis and using DT-based classification mining method to analyze the data in the student's sports achievement database, and generate the student's sports achievement DT, which can intuitively show the position of a certain achievement in different grade calculation methods and provide evaluation information for teaching departments [2]. Wu et al. used the classical C4.5 algorithm to construct the prediction model of students' professional

achievement, including data processing, pruning optimization of DT, and evaluation of the performance of the model by cross validation, to find out the potential students' behavior rules [3]. Yuan et al., based on the theory of performance evaluation, designed a performance evaluation index system to measure students' growth according to the process performance, taking innovative and designed experiments in the teaching stage of students' professional courses as the process [4].

Big data has the characteristics of large amount of data, various types, low-value density, high speed, and high timeliness. The era of big data poses new challenges to people's existing data control ability and also provides unprecedented space and potential for people to gain deeper and comprehensive insight. With the continuous accumulation and expansion of data, data mining is particularly important. Therefore, how to scientifically and reasonably mine students' sports achievement data and find the potential knowledge and information in the achievement data so as to guide students and teaching staff in the next stage of study and work has a positive guiding significance for the improvement of teaching quality. Therefore, this paper proposes to use RF (Random Forest) algorithm to accurately show many aspects of college students' sports performance analysis from a quantitative point of view so as to help teachers, students, and physical education departments formulate corresponding measures, which is conducive to the cultivation of students' physical exercise habits, the promotion of healthy physique, and the improvement of physical education teaching quality.

## 2. Related Work

*2.1. Analysis and Research on Sports Effect.* When human beings are engaged in sports activities, they cannot live without the environment, so they must exchange material, energy, and information with the natural and social environment. The factors that make up the environment are bound to be connected with sports activities as a whole or with each other.

Yi and Fan think that the school sports environment refers to the sports environment within a school, and the school sports environment, school moral education, learning, and other environments together constitute the school education environment [5]. Ouyang and Liu divided the campus sports environment into five subenvironments from the content, namely, sports consciousness environment, training environment, teaching environment, hygiene environment, and competition environment, and discussed the influence of subenvironments on students' physical and mental development [6]. Bideau et al., under the guidance of the concept of lifelong physical education, analyzed the result evaluation and the process evaluation so that the university physical education teaching evaluation can be evaluated from two aspects: process and result, making it more scientific and reasonable [7]. Li et al. calculated the weights of each index and each subsystem by AHP (analytic hierarchy process) and made single ranking and total ranking. The weighted linear method is used to synthesize

the indexes of the system to evaluate the development level of each subsystem and the system [8].

Yin and Cui studied the application of PCA (principal component analysis) and Bayesian KNN (*K*-nearest neighbor) algorithm in DM (data mining). PCA was used to reduce the dimension of characteristic data, and Bayesian *K*-nearest neighbor algorithm was used to sell and classify, that is, to predict career direction [9]. Sarlis and Tjortjis put forward an improved algorithm after researching and analyzing ID3 algorithm, which can mine and analyze the data stored in the educational administration management system, so as to find out the relationship between curriculum settings and provide some data basis for the university's achievement statistics to assist with decision-making [10]. Harvey et al. used model long-distance running to establish fuzzy evaluation model for data analysis. They solved the problem of conversion between national physical health standards and students' sports scores [11].

*2.2. Research on RF Algorithm.* RF is a multi-DT model that combines multiple DTs to make predictions. It has the advantages of ranking the importance of variables, high prediction accuracy, strong antinoise ability, few adjustment parameters, strong adaptability to data sets, and effectively avoiding "overfitting" phenomenon.

Poona et al. used RF to select the characteristic genes of five groups of gene expression profile data, such as lung cancer, stomach cancer, and colon cancer, and combined the selected characteristic genes with support vector machine to classify the original data set [12]. Provost et al. improved the node partition method of each tree in RF algorithm and proposed an improved RF algorithm. This algorithm takes the customer's life value as the dividing node and solves the problem of unbalanced data distribution [13]. Lee et al. used RF algorithm to classify different land types, and the results showed that RF algorithm could accurately classify land cover, which was consistent with expectations [14]. Naghibi et al. adjusted the weight of each sample according to whether the samples were correctly classified and proposed a weighted RF combining algorithm based on adaptive lifting [15]. Ming et al. applied greedy random adaptive search program with annealing randomness to RF algorithm, and its feature selection can improve the accuracy of classification [16].

Ghorbanian et al. designed and implemented an improved RF algorithm by using Spark distributed storage computing technology and proposed an improved RF algorithm based on Spark [17]. Bei et al. put forward the RF model of fusion factor analysis. The main innovation lies in constructing feature groups by factor analysis and then randomly extracting features according to the number ratio of features to form candidate subsets of each split node. Improve the accuracy and convergence speed of the model [18]. Raphael et al. used RF algorithm to study the nonlinear relationship between biomass pyrolysis kinetic parameters, biomass species, and various heating conditions and put forward an RF model to predict biomass pyrolysis kinetic parameters. The training results show that RF model can

predict biomass pyrolysis kinetic parameters under different heating conditions well [19]. Lin et al. used the improved RF algorithm to build an evaluation model based on the historical data of wheat breeding and put forward the application of RF algorithm in the auxiliary evaluation of wheat breeding [20]. The experimental results show that the proposed algorithm has achieved good results in wheat germplasm evaluation and can assist breeders in variety selection.

### 3. Research Method

**3.1. Analysis Method of Sports Effect.** Nowadays, the computer network has become very popular in universities. Using DM technology to establish a university sports achievement management system should be able to provide administrators, teachers, and students with sufficient information and quick query means, complete the work of teachers' scoring, and make statistics, analysis, and processing of data. It is conducive to scientifically set and arrange one's own fitness goals and exercise plans. Physical exercise can increase the opportunities of contact and communication between people. By taking part in sports activities, you can forget your troubles and pains, eliminate your loneliness, and gradually form the awareness and habit of interacting with others. It also helps students to change their bad behaviors and living habits that affect their health.

Cooperation ability is not only a necessary quality for participants in sports activities, but also a kind of ability that needs to be developed through sports activities. College students' participation in sports activities, especially in collective sports activities, requires the full cooperation between themselves and others, which not only enables the collective goal to be achieved, but also gives full play to the role of individuals.

DM is the process of extracting hidden information and knowledge from a large number of incomplete, noisy, fuzzy, and random data, which people do not know in advance but are potentially useful. It is of great significance to detect these anomalies and deviations, which can better optimize the model. There are many different types of deviations, such as special cases that do not conform to the basic rules, abnormal cases in clustering, values that change with time, and so on. From the perspective of DM, the application of DM technology in scientific management of physical examination results can be divided into three stages: data preparation, pattern search, and visual representation. Figure 1 shows the general framework model of scientific management of physical examination results.

The scientific management framework of physical examination results described in Figure 1 can be realized by four steps: the establishment of data objects, data preprocessing, data mining, and the representation and evaluation of pattern knowledge. In the scientific management of physical examination results, there are many data sets involved, such as teacher information, student information, examination results, and so on. Therefore, it is the first condition to establish a suitable DM object and make clear the goal of DM.

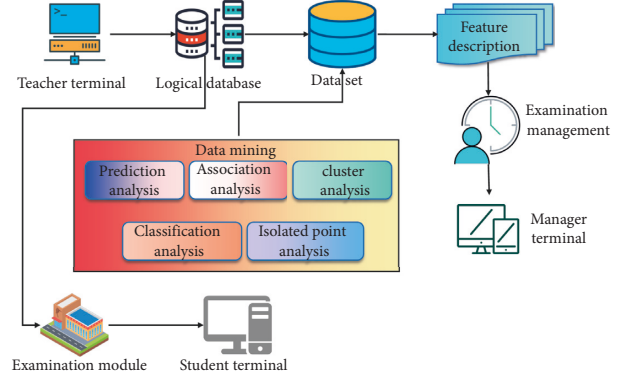


FIGURE 1: General framework of scientific management of general physical examination results.

Data preprocessing is an important link in the process of DM, especially when the data containing noise, incompleteness, or even inconsistency are DM; it is even more necessary to improve the quality of DM objects and finally achieve the purpose of improving the quality of pattern knowledge acquired by DM.

Information gain metrics tend to test with many outputs, that is, tend to select attributes with a large number of values [15]. Information gain rate is the expansion of information gain, which can overcome this problem. It normalizes the information gain by using the "split information" value. Split information is defined as follows:

$$\text{splitinfo}(A) = - \sum_{j=1}^v \frac{|D_j|}{|D|} * \log_2 \left( \frac{|D_j|}{|D|} \right). \quad (1)$$

This value represents the information generated by dividing the training data set  $D$  into  $v$  partitions corresponding to the attribute  $A$  test.

The submodel of RF algorithm is CART (Classification and Regression Tree) model. CART algorithm uses binary tree construction in binary recursion mode, and every time the whole sample set is divided into two subsets, thus generating two branches of subtrees. Gini index is mainly used to indicate the impurity of data set. Gini index of sample set  $D$  is defined as

$$\text{Gini}(D) = 1 - \sum_{i=1}^m p_i^2, \quad (2)$$

where  $p_i$  is expressed as the probability that the data in sample set  $D$  belongs to category  $C_j$ . If the binary partition of sample set  $D$  based on attribute  $A$  will be divided into two subsets  $D_1, D_2$ , then the Gini index based on this partition can be calculated as the weighted sum of the impurity of each partition:

$$\text{Gini}_A(D) = \frac{|D_1|}{|D|} \text{Gini}(D_1) + \frac{|D_2|}{|D|} \text{Gini}(D_2). \quad (3)$$

Applying DM technology to improve the level of systematic analysis, through in-depth and scientific analysis of examination results, students can know the relative position

of individuals in the group from horizontal and vertical aspects, make clear the gains and losses of individual learning and the differences with others, and promote individual learning. At the same time, teachers and school education decision-makers can understand the problems existing in teaching and reflect on the quality of education and teaching.

Introducing the observation set of sports effect evaluation, comprehensively analyzing and scheduling the sports effect evaluation model, and setting the value of the reliability weight  $\omega$  of sports effect evaluation are represented as

$$\omega = \omega_{\max} - t \frac{\omega_{\max} - \omega_{\min}}{T_{\max}}. \quad (4)$$

$\omega_{\max}$ ,  $\omega_{\min}$  represents the regulation coefficient of sports effect evaluation,  $T_{\max}$  is the time delay, and  $t$  is the correlation factor of sports effect evaluation. Combining with the scheduling method of association rules, the cloth fusion model of sports effect evaluation is obtained [6], and the calculation formula is

$$SL_i = \begin{cases} L_i, & i = 1, \\ \text{New}_i, & \text{otherwise.} \end{cases} \quad (5)$$

$\text{New}_i$  represents the distributed scheduling set of sports effect evaluation, thus constructing the fuzzy association rule set and prior distribution set of sports effect evaluation. By using the big data fusion analysis method [7], the characteristic analysis model of sports effect evaluation is established to improve the effectiveness of sports effect evaluation.

**3.2. Construction of Evaluation Model of Students' Sports Performance.** The construction of university sports environment indicators must follow the scientific principle, and the selection and construction of evaluation indicators should be based on scientific nature so as to avoid mistakes caused by empiricism and subjective judgment as far as possible so as to ensure the reliability of research results. Therefore, we should select the indicators from multiple dimensions, comprehensively and systematically, and try our best to make the selected indicators more comprehensive and systematic. Unless there are related scientific methods for conversion and measurement, it will affect the credibility of the evaluation. Ensure the operability, availability, and practicability of the evaluation process of college students' campus sports environment, which are the basis of accurate and feasible evaluation results and popularization and application.

Weight is a kind of quantitative value that compares and balances the evaluation object with a quantitative relationship to show its relative importance. According to different evaluation purposes, the weight set according to the importance of indicators is called the weight coefficient [17]. The general subjective weighting methods are illegal, analytic hierarchy process and weighting method. Subjective weighting requires evaluators to score the importance of each index and make an evaluation. Its characteristic is that it can humanize the opinions of experts and evaluate the importance of each index [19].

Although the accuracy and efficiency of the traditional RF algorithm are better than other classification algorithms in dealing with high-dimensional data, there are some defects in some data: if the data dealing with classification problems are unbalanced data, the classification results predicted by RF algorithm are not ideal; if the voting weights are the same, there will always be errors in the final results, especially on the unbalanced data sets, and the classification results tend to be biased towards most categories. The default parameters are not necessarily the best choice, and it is not efficient to use the traditional grid search method to find the best parameters.

In view of the above problems existing in the evaluation of students' sports performance, this paper puts forward an evaluation algorithm of students' sports performance based on RF algorithm, which can be used to evaluate students' sports performance, predict the students' related sports test scores of senior grades through the students' sports test scores of junior grades, sort the variables in importance, and find out the key factors that affect students' academic performance.

On the basis of the previous research, this paper proposes a new comprehensive optimization algorithm, which mainly integrates simulated annealing algorithm into the execution process of RF algorithm and uses binary coding,  $\text{OOB}_{(\text{OutofBag})}$  error minimization, and simulated annealing operation to obtain the best combination of feature selection  $O$ , DT scale  $K$ , feature subset scale  $N$ , and DT weight  $W_t$ .

We set the objective function of the improved RF algorithm as follows:

$$f \left( \begin{matrix} K^*, O^*, \{\text{Attribute}_i | i = 1, 2, \dots, M\}, \\ \{w_j | j = 1, 2, \dots, K\} \end{matrix} \right) \quad (6)$$

$$= \arg \min (\text{avg OOB error}),$$

where  $K, O$  is a real number,  $K \in [0, 500]$ , and  $w_j \in [0, 15]$ . The value of  $\text{Attribute}_i$  is 0 or 1, where 0 indicates that the feature has not been selected, and 1 indicates that the feature has been selected.

If the sample set is  $F$ , a few classes in the data set are  $F_{\min}$ , most classes are  $F_{\max}$ , and the imbalance coefficient  $E$  is shown in the following formula:

$$E = \frac{F_{\min}}{F_{\max}}. \quad (7)$$

The classification ability of each DT in RF is different. In the traditional RF algorithm, the votes of each DT are the same, which is unreasonable. In order to improve the classification ability of RF algorithm on unbalanced data, avoid the situation where the classification results of RF algorithm are seriously biased towards the majority class and make it perform well on the minority class of unbalanced data sets. The weight of DT voting is shown in the following formula:

$$W_i = E_i + \frac{\min_i}{L}. \quad (8)$$

$E_i$  is the unbalance coefficient of the data sampled from the training set of the  $i$ th DT,  $\min_i$  is the number of minority samples with correct prediction of this DT, and  $L$  is the length of data samples to construct this DT.

According to the Gini index, the importance of these factors is ranked, in order to help students make targeted tutoring according to the experimental results, guide the teaching work of teaching staff, provide decision-making support for education management departments, and improve teaching quality. Sample collection of students' sports achievements:

$$SC = \{(x_i, y_i), \quad i = 1, 2, \dots, n\}. \quad (9)$$

$X = X(X_1, \dots, X_n)$  is the explanatory variable, that is, the course name related to the course to be predicted, and  $Y$  is the target variable, that is, the student's sports achievement to be predicted.

The optimized prediction value of sports effect evaluation is as follows:

$$x(t_{n+1}) = X_{m+1}(m). \quad (10)$$

The structure of student sports performance evaluation model based on RF algorithm is shown in Figure 2.

The basic idea of RF algorithm is as follows:

- (1) From the original data set of students' sports scores after the above data preprocessing, the Bootstrap sampling method is adopted to randomly extract  $N$  scores training sets.
- (2) DT is established for each performance training set, and according to the principle of minimum Gini index, the best attribute is selected for internal node splitting.
- (3) Collect the prediction results of  $N$  DTs, and decide the grade category of the new sample according to the voting method.

## 4. Result Analysis

In order to verify the effectiveness of the design model in sports effect evaluation, MATLAB is used for simulation analysis, and SPSS statistical analysis software is used for big data analysis of sports effect evaluation. The distribution sample length of statistical data of sports effect evaluation is 1026, the frequency of feature sampling is 1.44 kHz, and the training set size is 300. The results are shown in Figures 3 and 4.

It can be seen that the classification ability of this algorithm is improved compared with that of RF algorithm. The same AUC (Area Under Curve) value under the default parameters of this algorithm and the AUC value after the optimization of RF algorithm parameters indicate the effectiveness of the improved algorithm. From the comparison of parameters, it can be seen that parameter optimization on data sets is helpful to improve the classification ability of the algorithm. The smaller the RF scale, the lower the depth of the tree, and the higher the efficiency, the better the

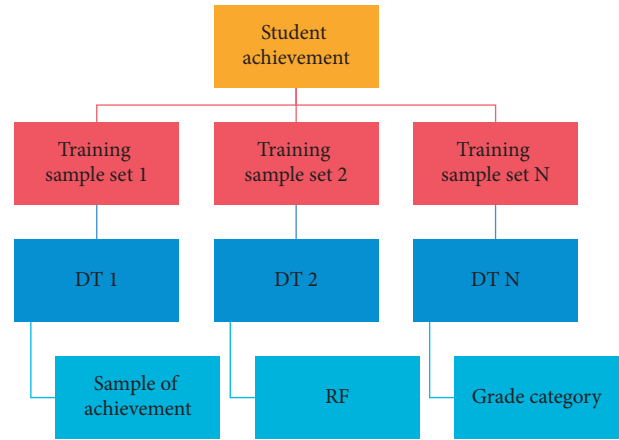


FIGURE 2: Evaluation model of students' sports performance based on RF algorithm.

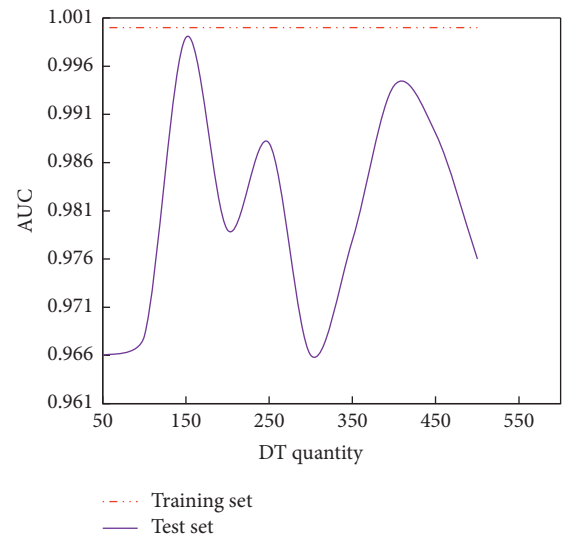


FIGURE 3: Influence of DT number on data set.

algorithm, under the condition that the Kappa coefficient and AUC value of the algorithm are guaranteed.

The quality of data determines the result of classification, so the data is preprocessed and then classified by RF. It can effectively improve the sensitivity of RF to unbalanced data. It is an important stage of college students' personal development, and all physical skills are developing vigorously. Excessive learning pressure may lead to students' physical and mental exhaustion. Therefore, we encourage students to participate in leisure sports. Moderate leisure sports can increase their vital capacity, promote students' quick thinking and clear mind, and help improve their learning efficiency. Through in-depth analysis of the existing achievements of students of the same major in universities, this paper explores the potential rules of achievement data, samples a large number of achievement data to form different training sets, and then forms a number of DTs to predict the students' achievements, respectively. Finally, the

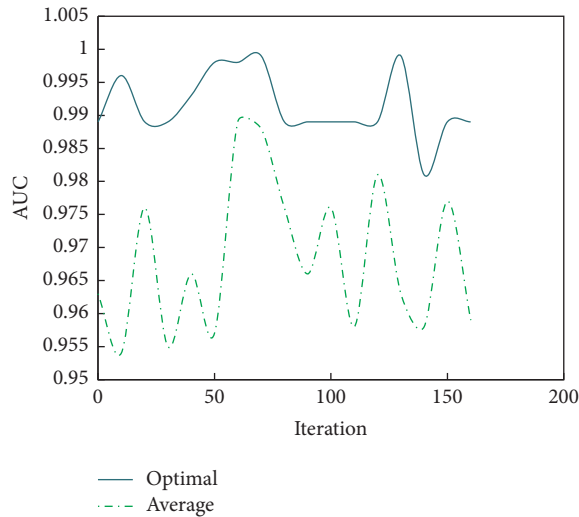


FIGURE 4: RF model iteration diagram.

risk level of students' sports achievements is obtained by integrating all DTs' prediction results.

Generally speaking, in independent tasks, competition has advantages, because in such tasks, the requirements for mutual cooperation among members are not very high, and the goal of individual activities is not to defeat others, but to accomplish the task. With the increasingly fierce competition in modern society, it is helpful for college students to adapt to the society well after they step out of school and enter the society by cultivating their competitive consciousness and ability. In order to ensure the fairness of the evaluation of college sports, it is convenient and accurate to use unified evaluation methods and means in the evaluation process, to make the subjective evaluation content tangible and digital, and to use computer data processing software for evaluation. In addition, it is necessary to combine statistical evaluation with expert evaluation to establish a fair and just evaluation system.

For the data set of students' sports achievements, the feature variables that need feature planning include students' activity in class, students' absenteeism, the number of times that students visit teaching resources after class, the number of times that students participate in course discussions, and students' satisfaction with the course. Other feature variables are calculated by using original values, such as birthplace with coded values and gender with Boolean values of 0 or 1.

In the process of realization, firstly, the discrete data is normalized to  $[0, 1]$  based on the log function standardization method, and the corresponding weight coefficient is assigned to each factor so that the students' classroom activity data can be weighted. Figure 5 shows the comparison of the classification results of students' sports performance data sets on various basic algorithms.

Figure 5 shows that we use a variety of basic DM algorithms for comparison and classification, trying to analyze the differences of various DM algorithms. From the experimental results, it can be seen that the RF algorithm has a

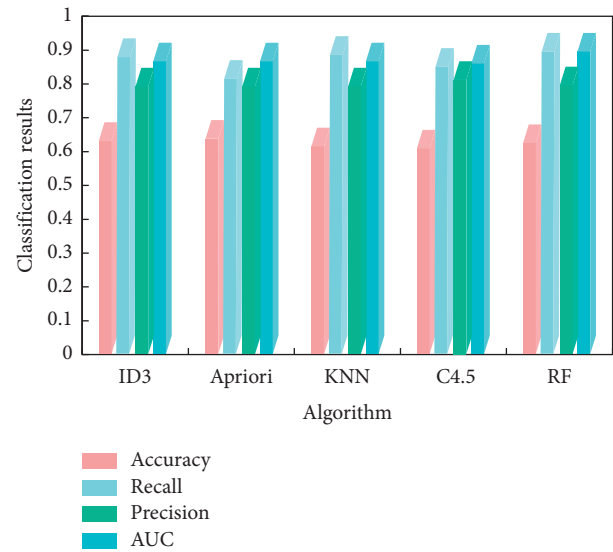


FIGURE 5: Comparison of classification results of students' sports performance data set on various basic algorithms.

TABLE 1: Training data.

Project	Passed number	Number of failed students
Long jump	80	77
Long-distance run	55	102
Basketball	30	122
Short-distance run	28	126
Hurdle race	39	115
Volleyball	97	79
Shot	36	120
High jump	99	59

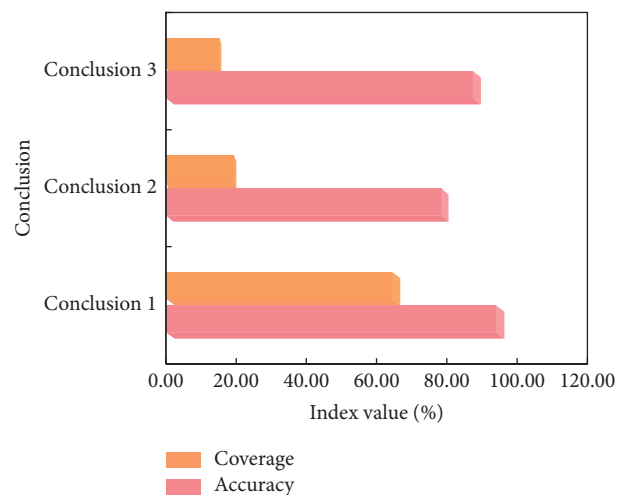


FIGURE 6: Schematic diagram of conclusion.

slight advantage over other data algorithms in the student characteristic data set, but the overall difference is not big.

TABLE 2: Statistical analysis of sports effect evaluation.

Group	Sample number	Standard deviation	Maximum	Minimum
Contrast group	286	0.533	33.68	12.96
Training group	632	0.306	32.17	15.62

What attributes can determine whether students can pass the sprint by using DT technology? Take the whole sampling method to select some students' sports achievements as training sets, and copy these achievements records to the training example worksheet. Using the function, find out the number of people who passed and failed in a single subject in the training set, as shown in data Table 1.

The key to construct a good DT lies in how to choose a good logical judgment or attribute. It has been found that, in general or with a high probability, the smaller the tree, the stronger the prediction ability of the tree. To construct DT as small as possible, the key lies in choosing appropriate logical judgment or attribute. Here, the information gain is used for attribute selection.

According to the information gain of each attribute, the attribute of long-distance running should be selected as the root node of DT. Other results are used as the test set and used to test the accuracy of the generated DT. The following rules can be obtained from this DT, as shown in Figure 6.

The learning level of students' long-distance running will directly affect their learning effect of sprint events. The learning of hurdle events also has a certain influence on the learning of sprint events. Therefore, teachers should consider students' long-distance running foundation when teaching sprint. Students with good long-distance running and average sprint should pay more attention to the study of hurdle events.

Compared with the culture course, physical education class is not valued by many schools and parents, and because of the pressure of entering a higher school, the duration of students' physical exercise is shortened. Parents cannot fully recognize the benefits of physical activities for students and attach too much importance to the achievements of culture course, which leads to their neglect of the cultivation of students' physical quality. Due to the negative effects on the body, leisure and psychology after taking part in physical exercise, such as physical fatigue after long exercise time or high intensity, learning collapse will be affected. In addition, under the pressure of traditional ideas and crazy social environment, physical exercise and academic performance are negatively correlated in form.

By comparing the factors that affect students' physical exercise at each level, we can find that the experience of individual participation in physical education class or physical exercise constitutes the common factors that affect students' participation in physical exercise. In behavioral learning theory, operant conditioning holds that learning is based on the response results. Poor students in physical education class: some students are often criticized in class because they do not obey discipline or behave badly. In this case, even if they have their own favorite sports, their performance in physical education class is restricted, and

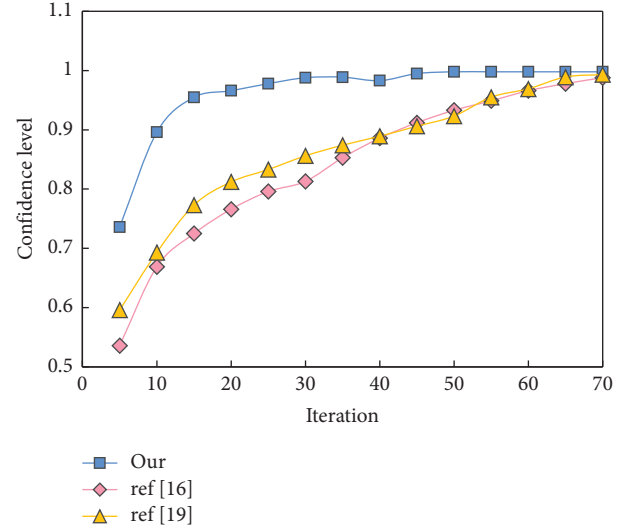


FIGURE 7: Comparison of confidence degree of sports effect evaluation.

their sports performance is not very good, which affects their understanding of sports to a certain extent.

Some students, though not doing well in the cultural course, have been active in physical education class, received positive responses, and had a positive experience in physical education class. According to the above simulation parameters, the sports effect is evaluated, and the statistical distribution sample set is shown in Table 2.

According to the results in Table 2, the big data analysis of sports effect evaluation is carried out, and the big DM results of sports effect evaluation are shown in Figure 7.

According to the analysis of Figure 7, the confidence level of sports effect evaluation by this method is high, and the average confidence level is above 0.96. The results of these professional practice courses are mainly graded by teachers according to the experimental results made by students. After observing the data, it is found that the scores of these courses are concentrated compared with those of professional theoretical courses, and the difference of scores is not obvious, which has little effect on analyzing students' later professional courses. Therefore, the experimental results are consistent with the actual situation. Among many factors that affect students' poor physical performance, "exercise situation" has the greatest influence on students' performance. For a student who never exercises, if his attendance rate in class is low and his foundation is average, then his exam results must be poor.

College students play an important role among the members of the society and undertake a strong social mission. Whether one can fulfill one's own responsibilities and gain good trust from others in sports not only plays an

important role in promoting one's role in sports, but also makes certain adaptation for college students to enter the society in advance. Strengthen the education of college students' safety knowledge, improve their self-management ability, formulate a reasonable and planned extracurricular physical exercise schedule, and select sports with low sports risks characterized by fun, safety, health, and positivity according to the value of the risk factors studied so as to reduce the occurrence of risks.

Physical education teachers should constantly optimize the educational methods. Teachers should learn and progress from each other, probe into each other, sum up experiences, form an educational method suitable for college students, and guide college students to love sports, cultivate their interest in sports, and participate in sports consciously. Set up the leading organization of extracurricular sports together, formulate corresponding rules and regulations, and accomplish all tasks in organizing extracurricular sports activities in schools. In organizing extracurricular sports activities, more college students have the opportunity to participate in sports, feel the charm of sports, and experience the fun of sports so as to participate in sports exercises and improve their physical quality.

## 5. Conclusion

Physical education is an important part of school teaching. How to make computer and network technology serve the daily physical education teaching in schools is an inevitable requirement under today's big data. Based on RF algorithm, this paper puts forward a scientific evaluation algorithm for students' sports achievements. Through this research, combining pattern recognition and big data analysis methods, the optimization of sports effect evaluation is realized. The research shows that this method has a high confidence level in sports effect evaluation, and the evaluation results are accurate and reliable. It can correct and intervene students' learning in advance so that students can achieve better results in professional courses and play a very good scientific guiding role in assisting students' learning behaviors and optimizing educational decisions.

This study only considers the relationship between participation in physical exercise and academic performance from the perspective of a school. As the author's ability is limited, this study does not take into account factors such as school differences, which is also something that needs to be explored in future studies.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## References

- [1] F. Yang and F. W. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining," *Computers & Education*, vol. 123, pp. 97–108, 2018.
- [2] W. S. A. A. Attar, N. Soomro, P. J. Sinclair, E. Pappas, and R. H. Sanders, "P-7 Effect of injury prevention programs that include the nordic hamstring exercise on hamstring injury rates in soccer players: a systematic review with meta-analysis," *British Journal of Sports Medicine*, vol. 50, no. Suppl 1, pp. A35.2–A35, 2016.
- [3] Z. Wu and C. Zhou, "Construction of an intelligent processing platform for equestrian event information based on data fusion and data mining," *Journal of Sensors*, vol. 2021, no. 6, 9 pages, Article ID 1869281, 2021.
- [4] C. Yuan and Y. L. YYang, "Sports decision-making model based on data mining and neural network," *Neural Computing & Applications*, no. 2, pp. 1–14, 2020.
- [5] D. Yi and G. Fan, "Sports video mining via multichannel segmental hidden markov models," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1301–1309, 2009.
- [6] J. Q. Ouyang and R. Liu, "Ontology reasoning scheme for constructing meaningful sports video summarisation," *IET Image Processing*, vol. 7, no. 4, pp. 324–334, 2013.
- [7] B. Bideau, F. Multon, R. Kulpa et al., "Virtual reality applied to sports: do handball goalkeepers react realistically to simulated synthetic opponents?" *Journal of Multivalued Logic and Soft Computing*, vol. 18, no. 5–6, pp. 228–233, 2012.
- [8] Y. Li, J. Ge, and W. Hao, "Construction and simulation of a multiattribute training data mining model for basketball players based on big data," *Wireless Communications and Mobile Computing*, vol. 2021, no. 7, 14 pages, Article ID 6399266, 2021.
- [9] Z. Yin and W. Cui, "Outlier data mining model for sports data analysis," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 1–10, 2020.
- [10] V. Sarlis and C. Tjortjis, *Information Systems*, vol. 93, Article ID 101562, 2020.
- [11] S. Harvey, O. Atkinson, and B. P. Hyndman, "Chapter 7: an investigation into sports coaches' twitter use," *Journal of Teaching in Physical Education*, vol. 39, no. 4, pp. 481–490, 2020.
- [12] N. K. Poona, A. Van Niekerk, R. L. Nadel, and R. Ismail, "Random forest (RF) wrappers for waveband selection and classification of hyperspectral data," *Applied Spectroscopy*, vol. 70, no. 2, pp. 322–333, 2016.
- [13] F. Provost, C. Hibert, and J. P. Malet, "Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier," *Geophysical Research Letters*, vol. 44, no. 1, pp. 113–120, 2017.
- [14] K. Lee, M. Lee, and D. Kim, "Utilizing random Forest QSAR models with optimized parameters for target identification and its application to target-fishing server," *BMC Bioinformatics*, vol. 18, no. S16, p. 567, 2017.
- [15] S. A. Naghibi, K. Ahmadi, and A. Daneshi, "Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping," *Water Resources Management*, vol. 31, no. 9, pp. 2761–2775, 2017.
- [16] D. Ming, T. Zhou, M. Wang, and T. Tan, "Land cover classification using random forest with genetic algorithm-based parameter optimization," *Journal of Applied Remote Sensing*, vol. 10, no. 3, Article ID 035021, 2016.



- [17] A. Ghorbanian, S. Zaghian, R. M. Asiyabi, M. Amani, A. Mohammadzadeh, and S. Jamali, "Mangrove ecosystem mapping using sentinel-1 and sentinel-2 satellite images and random forest algorithm in google earth engine," *Remote Sensing*, vol. 13, no. 13, p. 2565, 2021.
- [18] Z. Bei, Z. Yu, N. Luo, C. Jiang, C. Xu, and S. Feng, "Configuring in-memory cluster computing using random forest," *Future Generation Computer Systems*, vol. 79, pp. 1–15, 2018.
- [19] C. Raphael, P. Philipp, and B. Anne-Laure, "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC Bioinformatics*, vol. 19, no. 1, p. 270, 2018.
- [20] W. Lin, W. Fan, H. Liu, Y. Xu, and J. Wu, "Classification of handheld laser scanning tree point cloud based on different KNN algorithms and random forest algorithm," *Forests*, vol. 12, no. 3, p. 292, 2021.