

## Retraction

# Retracted: Application of Data Mining Technology in Public Welfare Sports Education in the Era of Artificial Intelligence

### Mobile Information Systems

Received 31 October 2023; Accepted 31 October 2023; Published 1 November 2023

Copyright © 2023 Mobile Information Systems. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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) Peer-review manipulation

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] X. Dong, X. Huang, and M. Lin, "Application of Data Mining Technology in Public Welfare Sports Education in the Era of Artificial Intelligence," *Mobile Information Systems*, vol. 2021, Article ID 8692292, 13 pages, 2021.

## Research Article

# Application of Data Mining Technology in Public Welfare Sports Education in the Era of Artificial Intelligence

Xiaoyan Dong,<sup>1</sup> Xiaohua Huang,<sup>2,3</sup> and Min Lin <sup>4</sup>

<sup>1</sup>Qixin College, Ningbo Tech University, Ningbo 315100, China

<sup>2</sup>School of Design, Ningbo Tech University, Ningbo 315100, China

<sup>3</sup>Faculty of Innovation and Design, City University of Macau, Macau 999078, China

<sup>4</sup>Physical Education, Hangzhou DingZi University, Hangzhou 310018, China

Correspondence should be addressed to Min Lin; 631418040103@mails.cqjtu.edu.cn

Received 6 August 2021; Revised 18 September 2021; Accepted 22 October 2021; Published 8 November 2021

Academic Editor: Omar Cheikhrouhou

Copyright © 2021 Xiaoyan Dong 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.

Data mining refers to extracting the implicit prediction information from a massive dataset. It has very application prospects. Some data mining tools can develop things. The purpose of this article mainly discusses the public welfare sports education in the artificial intelligence era. The article discusses the research background and significance, development of education data mining, and decision tree technology and enumerates the application of education data mining in real life. The concept of educational data mining is given, and several common typical decision tree algorithms and their connections and differences are described; then, the concepts of multivalued decision tables and decision trees are discussed in detail. This article aims to build a nonprofit physical education system to manage and analyze the attendance data of students' physical health assessment, so as to improve the enthusiasm of students to exercise, such as BP neural network, decision tree classification algorithm, and cluster analysis, discusses the calculation and analysis process of the relevant body side data of the sports teaching platform, and emphatically discusses and analyzes the application effect of data mining technology in public welfare sports teaching. In addition, this article has built a public welfare physical education system, allowing us to clearly understand various factors that affect students' exercise and the relationship between various project indicators. Based on these data, educators can adjust technical means. Experimental results can efficiently and conveniently understand the pass rate of students in various sports. The pass rates of students in the six tests, grip strength, and sitting forward bending were 58%, 65%, 78%, 78%, 85%, and 65%, respectively. Using mathematical methods and computer technology, we can dig out valuable education management information from massive education data, so as to provide a reference for improving school enrollment.

## 1. Introduction

**1.1. Background.** For storing massive amounts of data, database technology was invented, but when the database was just born, people could only store the existing data in this way because database technology did not have the function of data analysis. In order to analyze data more deeply, understand the meaning contained in these data, and find out the correlation between data and data, data mining technology came into being. At the same time, at the beginning of the 21st century, our country has carried out a comprehensive reform of basic education. New ideas, new theories, and new methods have not only put forward new requirements and goals for primary and secondary education but also put forward new goals and

challenges for teachers, especially physical education teachers. As the scale of the development of public welfare sports becomes larger and larger, more information is stored in its sports system, including attendance information and physical fitness assessment data, and then it becomes an inevitable requirement management. Moreover, manual management of information is prone to errors, and students' enthusiasm for physical exercise will be affected. Therefore, how to organize sports data conveniently and quickly with low error rate and high efficiency is a problem to be solved immediately [1].

**1.2. Significance.** The technology of deriving hidden prediction information from large databases or data storage is a very popular data mining technology in today's society. It

can not only help people find hidden data patterns but also find the most valuable information and knowledge. The database server can automatically analyze and extract data according to various data mining techniques of the driver; in addition, people do not need to worry about the complex calculation and complexity and type and process of the basic data mining method of the application because it can simply compose a complex analysis model by itself. The database server can easily and quickly extract massive amounts of data, such as student fitness test data, to provide decision-makers with useful information, make more rational use of the school's limited resources, and create a higher education environment. Therefore, we implemented and studied data mining technology in the course. The physical education system strengthens the teaching ability of management personnel to process. The sport management system further optimises and enriches the data resources. At the same time, mining the data of various aspects of the student's physical condition can find the evaluation results of the student's physical condition and the correlation between different data. To discover the problems of students in physical exercise in time, administrators and educators can use these data to monitor the physical condition of students [2].

*1.3. Related Work.* At present, our country's education system has developed to a certain stage. As the country's development enters a new normal, educational structural contradictions have become increasingly prominent. As a part of the higher education system, physical education is also facing new opportunities and challenges, and the sports talents cultivated by the previous sports education model can no longer meet the requirements. Zhang and Li mainly studied employment and entrepreneurship data of college graduates over the years and mined employment and entrepreneurship data. The authors established a data mining model for data mining analysis with the help of Weka tools and applied the improved ID3 algorithm to the data mining model. The research has drawn the conclusion that male and female graduates have different employment goals [3]. Liang analyzed the application of political education of vocational students. Finally, on this basis, a countermeasure was proposed to optimise the education policy in colleges and universities [4]. Li et al. proposed to store the valid data and the original data in the database, so as to facilitate the teaching and research staff to compare relevant parameters and make decisions. To a certain extent, the data is calculated by cluster analysis technology algorithm [5]. Tao et al. briefly introduced the theoretical basis of data mining technology, including definition, research content, essence, and function. Through the design of data mining functions and development models, the final conclusion is that learning data mining technology can effectively improve the level of information management and provide effective management methods for university managers [6]. Hemantkumar and Adholiya proposed the use of educational data mining to improve the performance of graduate students. The authors studied educational data by applying different techniques such as neural networks,

association rules, regression, Bayesian networks, and rule-based systems. Association rule mining technology can discover the relationship between variables in large databases and finally achieve the purpose of classifying students according to their academic performance on this basis [7]. Lu and Zhou wrote this article to focus on the development of applied technology for the extraction of mathematical education. Second, use algorithms to extract information and classify and integrate a large number of trained professionals [8]. Although the analysis is very accurate, there are some shortcomings in the analysis of physical education.

*1.4. Innovation.* The innovations of this article are as follows: (1) the first is the innovation of the topic selection angle. This article is a new perspective from the perspective of topic selection. At present, there are not many research studies that integrate artificial intelligence, data mining, public welfare sports, decision tree algorithms, clustering, and neural networks. It is of exploratory significance. (2) The second is the innovation of research methods. Various data mining methods such as the BP neural network, decision tree algorithm, and cluster analysis algorithm are proposed, which have theoretical value. (3) The other is the innovation of project practice. The project has formed an efficient and complete non-profit physical education system, so as to discover students' problems in physical exercise in time. Besides, administrators and teachers can use these data to monitor the physical condition of students.

## 2. Data Mining Technology

*2.1. Algorithm of the BP Neural Network.* The output is preexisting [9]. The algorithm consists of two parts: the direct transmission process of the input information and the weight adjustment error [10] backpropagation process. In the process of direct transmission, information is transmitted between neurons. If there is no expected value in the output, it will be propagated backward, and the weight will be adjusted by returning the error signal to the original path. Assume that the is  $x \in R^n$ , of which  $x = (x_1, x_2, \dots, x_n)^T$ .

- (1) The output of the  $i$ th neuron in the hidden layer is expressed as

$$s1_i = r1 \left( \sum_{j=1}^{j=n} w1_{ij} x_j + b1_i \right), \quad i = 1, 2, \dots, s1. \quad (1)$$

- (2) The output of the  $g$ th neuron in the output layer is expressed as

$$s2_k = g2 \left( \sum_{i=1}^{s1} w2_{ki} d1_i + b2_k \right), \quad k = 1, 2, \dots, s2. \quad (2)$$

- (3) The error function is defined as

$$T(H, B) = \frac{1}{2} \sum_{k=1}^{s_2} (t_k - a_{2k})^2. \quad (3)$$

Changes in the weights of the output layer are

$$\Delta h_{2kj} = -\eta \frac{\partial E}{\partial w_{2kj}} = -\eta \frac{\partial E}{\partial a_{2v}} \cdot \frac{\partial a_{2k}}{\partial v_{2kj}} = -\eta (t_k - a_{2k}) \cdot f_{3sxc}. \quad (4)$$

Changes in the threshold of the output layer are

$$\Delta h_{1ik} = -\eta \frac{\partial E}{\partial b_{3kj}} = -\eta \frac{\partial E}{\partial x_{2k}} \cdot \frac{\partial a_{2k}}{\partial b_{1ki}} = -\eta (y_k - a_{2k}) \cdot l_{2'}. \quad (5)$$

Changes in hidden layer weights are

$$\Delta f_{kl} = -\eta \frac{\partial E}{\partial w_{lij}} = -\eta \frac{\partial E}{\partial d_{2k}} \cdot \frac{\partial a_{2k}}{\partial a_{sh}} \cdot \frac{\partial a_{li}}{\partial w_{lij}} = -\eta \sum_{k=1}^{s_2} (l_k - a_{2k}). \quad (6)$$

Changes in the hidden layer threshold are

$$\Delta b_{li} = \eta \sum_{k=1}^{s_2} (y_k - a_{2k}) \cdot f_{1'w1kj}. \quad (7)$$

**2.2. Decision Tree Classification Algorithm.** C4.5 has two main steps: first, build an appropriate data model according to the dataset and the meaning it wants to express [11]; second, perform data analysis. Figure 1 shows the basic principle of the decision tree algorithm.

An important notion in C4.5 algorithm is the speed of obtaining information [12], and the selection of the characteristics shall be made by calculating the information transfer factor of each characteristic when the decision tree model is created. Set the dataset as  $W$ , and the quantity of classification information can be expressed as [13]

$$\text{info}(W) = \sum_{i=1}^n f_i \log_2 f_i. \quad (8)$$

Suppose one of the attributes, marked as  $B$  here, has a different value, and these values can be expressed as  $\{B_1, B_2, \dots, B_n\}$ ; therefore, the attribute  $B$  can divide the dataset  $W$  into  $n$  different subsets, which are denoted as  $\{W_1, W_2, \dots, W_n\}$  here, assuming  $F_{ij}$  is set to  $A$  [14] is follows:

$$\text{info}_A(W) = \sum_{j=1}^n \frac{f_{1j} + f_{2j} + \dots + f_{mj}}{d} \text{info}(f_{1j}, f_{2j}, \dots, f_{mj}). \quad (9)$$

The formula for  $F_{ij}$  is  $(f_{1j} + f_{2j} + \dots + f_{mj}/d)$ . The information volume of subset  $W$  [15] is calculated as follows:

$$\text{info}(f_{1j} + f_{2j} + \dots + f_{mj}) = \sum_{i=1}^m P_{ij} \log_2(P_{ij}). \quad (10)$$

$p_{ij} = (f_{ij}/d_j)$  means that the samples in dataset  $W_j$  account for the proportion of class  $f_j$ .

Then, the functional expression of feature  $A$  is

$$\text{Oain}(A) = \text{info}(W) - \text{info}_A(W), \quad (11)$$

that is,  $\text{SplitInfo}_A(W)$  [16]:

$$\text{SplitInfo}_A(W) = \sum_{j=1}^n p_j \log_2 p_j, \quad (12)$$

and  $p_i$  indicates that the value of attribute  $A$  is the proportion of  $k_i$  in the total dataset [17], which can be calculated from  $f_j/f$ .

$$\text{GainRatio}(W) = \frac{\text{info}(W) - \text{info}(W)}{\text{SplitInfo}_A(W)}. \quad (13)$$

Finally, the above process is repeated to complete the calculation.

**2.3. Cluster Analysis Technology.** Group virtual or physical datasets, with the purpose of compiling datasets from similar data objects. These data object groups created by the cluster analysis process are called clusters [18], which are a set of data objects. The data mining technology [19] scheme is shown in Figure 2.

**2.3.1. Data Matrix.** For all  $k$  objects, select  $z$  variables to describe, measure these data types with the interval scale, and then express them in the form of a relational table, that is,  $k \times z$  matrix:

$$\begin{pmatrix} X_{11}, X_{12}, \dots, X_{1z} \\ X_{21}, X_{22}, \dots, X_{2z} \\ \dots \\ X_{k1}, X_{k2}, \dots, X_{kz} \end{pmatrix}. \quad (14)$$

**2.3.2. Dissimilarity Matrix.** The clustering of the Kohonen network [20] is based on the learning of self-organizing features; in this learning process, the principle of lateral interaction is used. In the learning process, a "cluster area" is formed near the neurons that win at each stage. The overall effect is to bring the weight vectors around these neurons close to the input vector values and finally collect input vectors with similar characteristics. The weight of the learned Kohonen network is sufficient to represent the characteristics of the input vector, so based on this, the input vector is classified and identified by the weight. The core is optimal neuron matching and weight self-organization distribution. At the same time, this selection process is the process of identifying the central neuron corresponding to the input pattern, where the specific quantification form of the dissimilarity between  $i$  and  $j$  is the two objects  $f(i, j)$ .

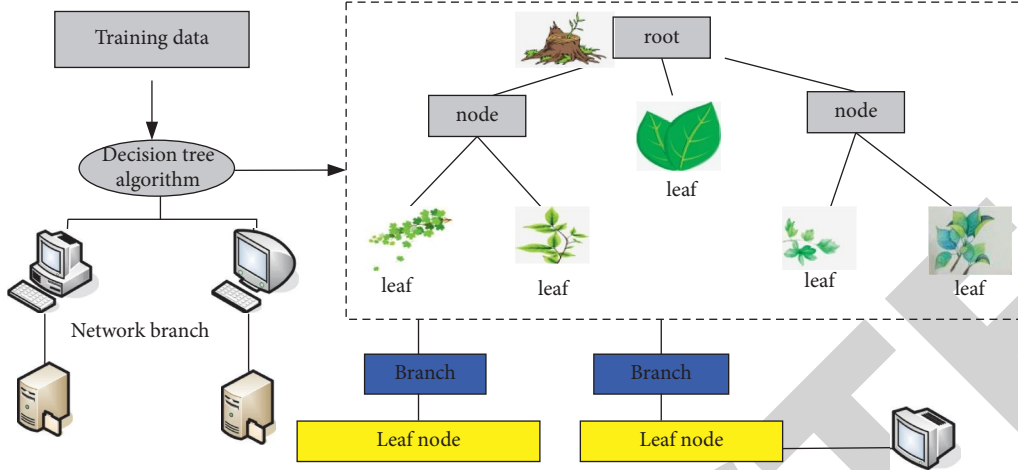


FIGURE 1: Basic principle of the decision tree algorithm.

$$\begin{pmatrix} 0 \\ f(2,1),0 \\ f(3,1),f(3,2),0 \\ \dots \\ f(n,1),f(n,2),\dots,0 \end{pmatrix}. \quad (15)$$

(1) *Interval-Scaled Variables*. The degree of [21] (or similarity) is usually measured by the distance between each group of objects. The most classic method of calculating distance is Euclidean distance, and researchers have summarized its definition.

$$f(i,j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{in} - x_{jn}|^2}. \quad (16)$$

Another very classic distance measurement method is Manhattan (or city block) distance [22]. The researchers summarized its definition as follows:

$$f(i,j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}|. \quad (17)$$

Minkowski distance (Minkowski) [23] is a generalization of the above two distances. The researchers summarized its definition as

$$f(i,j) = \|x_i - x_j\|_p = \left( \sum_{k=1}^n |x_{ik} - x_{jk}|^p \right)^{1/p}. \quad (18)$$

(2) *Binary Variables*. Each variable has two statements [24]. Symmetrical variables and asymmetrical variables are two types of binary variables. If the values contained in the binary variables of the two states of 0 and 1 are equally important, they will be called symmetrical binary variables. When researchers describe the degree of inconsistency between object  $i$  and object  $j$ , they usually use a simple

matching correlation coefficient to describe it, which can be defined as the following formula:

$$f(i,j) = \frac{(g+d)}{(q+c+d+t)}. \quad (19)$$

If the meanings included in the two states of 0 and 1 of a binary variable are not equally important, then the binary variable is said to be an asymmetric binary variable. It can be defined as the following formula:

$$f(i,j) = \frac{(g+d)}{(q+k+b)}. \quad (20)$$

(3) *Categorical Variables*. The value of categorical variables often has several more states than binary variables [25]. The degree of dissimilarity between categorical variables  $i$  and  $j$  can be used to calculate:

$$f(i,j) = \frac{p-m}{p}. \quad (21)$$

(4) *Ordinal Variables*. Ordinal variables are divided into discrete ordinal variables and continuous ordinal variables [26]. Since the number of states of each ordinal variable is not the same, it is usually necessary to direct the values of all variables to the area [0.1, 2.0]; the purpose is to make all variables have the same weight, which can be replaced by implementation, where

$$d_{kf} = \frac{g_{kf} - 1}{W_f - 1}. \quad (22)$$

If the variable  $f$  is an interval-scaled variable, then [27]

$$f_{if}^{(f)} = \frac{|x_{if} - x_{jf}|}{\max_v x_{fv} - \min_v x_{fv}}. \quad (23)$$

It is mean between the objects in each cluster, that is, the center of mass or center of gravity in the cluster. The

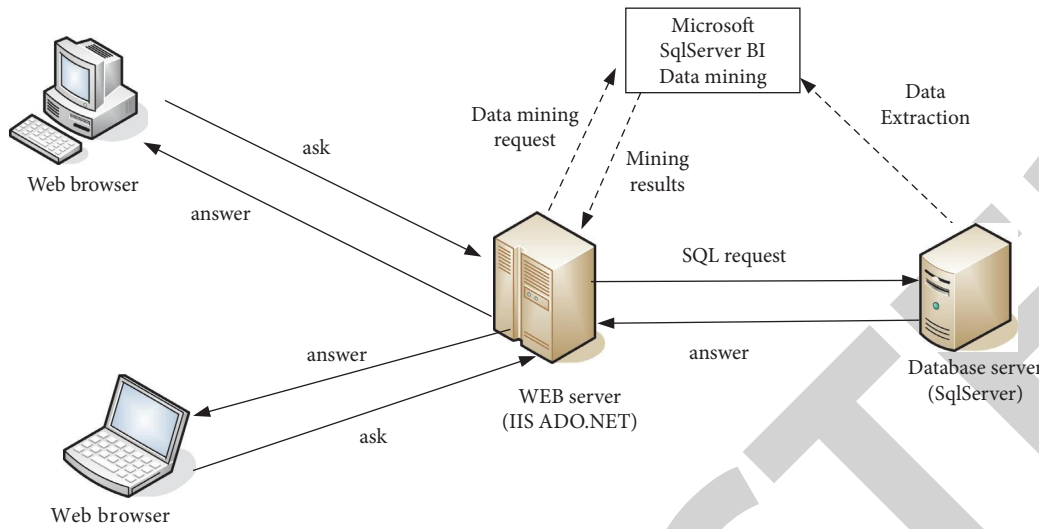


FIGURE 2: Mining data analysis program.

commonly used objective function is the quadratic error criterion function [28, 29]:

$$E = \sum_{i=1}^b \sum_{p \in c_i} |p - m_i|^2. \quad (24)$$

### 3. Construction and Design of the Public Welfare Sports System

**3.1. The General Structure of the Public Welfare Sports System.** The object of public welfare sports is all teachers and students, so that it is convenient for teachers to support the B/S model and the popular SSH platform. The network layer is the entrance to the public welfare sports system, which provides different services according to different users. This part is realized by ZK. For physical education teachers, students, system administrators, school leaders, and other users, the service layer is the most important part of the entire system. Business requirements such as business rules and business processes are reflected in this level. This class, also known as the logical class, is completed by spring. The service layer has the responsibility to accept the request and interact with basic data through pages and logic processing. The DAO layer, also known as the data persistence layer, can directly manipulate the database. The physical layer is the sports system database, which mainly includes user data, attendance data, and body measurement data. Both the DAO layer and the entity layer are supported by Hibernate. The general structure of the public welfare sports system is shown in Figure 3.

**3.2. Design of the Public Welfare Sports System.** According to the trial plan of the “Student Physical Health Standards” and the actual implementation of “Public Sports” in various universities, the public interest sports system can be divided into five aspects. Specific services include the personal center module, system management module,

sports attendance module, and help and feedback module. Among them, the personal center module includes personal information (my collection, personal information, and password modification) and message center (notification announcement, personal message, and news center); the system management module is maintained by the title (configure the resources and permissions owned by each role), role management (add, modify, and delete a role), organization management (maintain the information of schools, colleges, and majors), and user management (manage all user information and their online status). The personnel management module includes managing campus users (adding users to different roles) and managing class student cadres (adding, modifying, and deleting information). The sports attendance module has 5 modules: equipment management, personnel management, attendance management, school calendar management, and score management. The help and feedback module has three parts: user feedback (users add and delete feedback), manage feedback (view user feedback), and document assistance (helps users understand system features and how to use them). The system function module is shown in Figure 4.

#### 3.3. Data Preparation

**3.3.1. Data Collection.** The data collection mentioned here should include structured data and semistructured data. Then, the Data Transformation Server (DTS) mainly performs structured data collection (including databases and heterogeneous databases with the same structure as SQL Server), and semistructured data can be obtained through the FTP and HTTP transmission. This system mainly uses structured data. This paper collects test data on height, weight, vital capacity, gait test, and grip strength. All classes in the 2012–2015 academic year are tested for height, weight, vital capacity, step test, grip strength, sitting body propulsion, and standing long

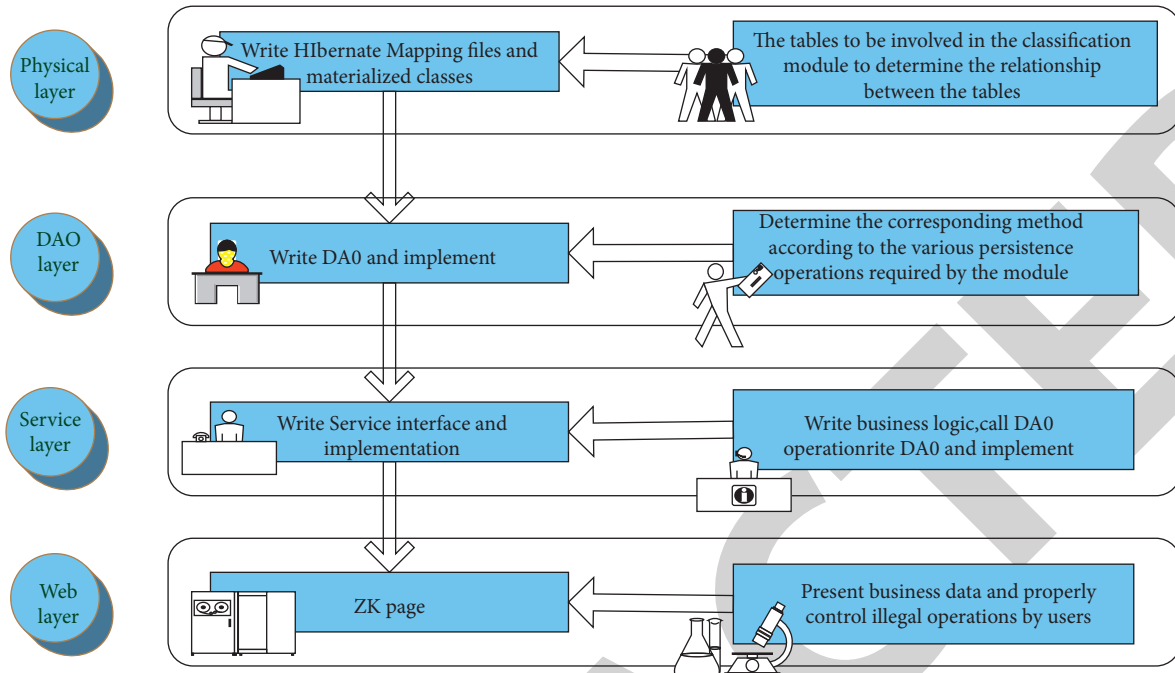


FIGURE 3: The overall structure of the public welfare sports system.

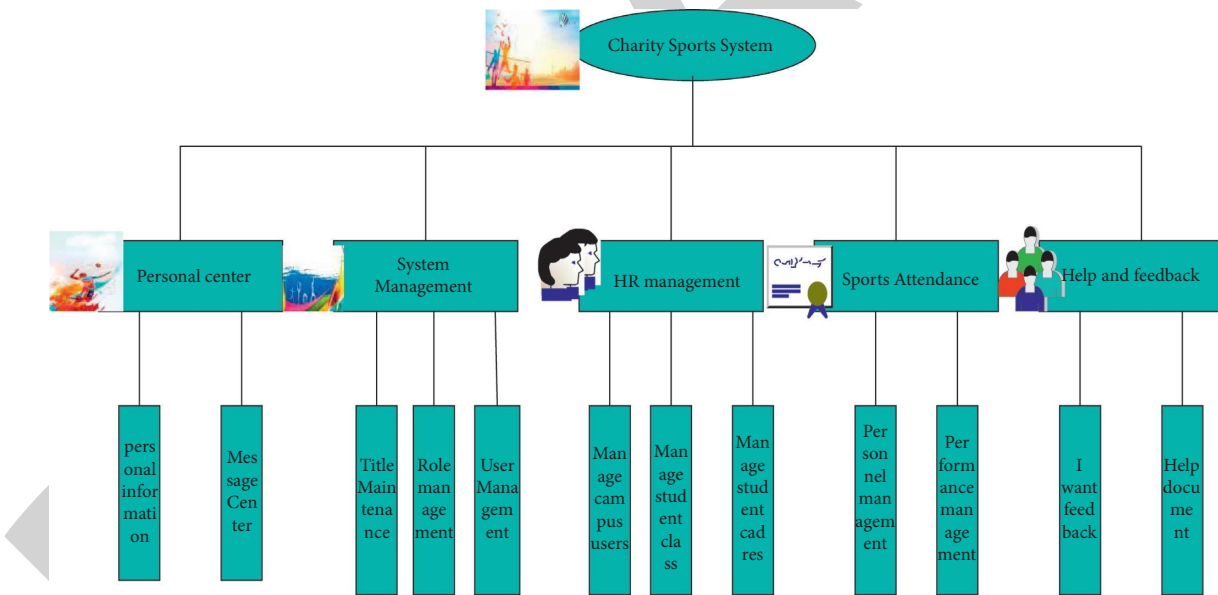


FIGURE 4: System function modules.

jump, and then database technology is used to create a data mining library for this article.

3.3.2. *Data Pretreatment.* Data preprocessing and its workload account for 70% of the entire process of data mining technology. The past practice is to save data modeling time after compiling data, greatly improving efficiency and accuracy. The data in this article include basic student information, attendance information, student health evaluation information, and other data. A large amount of data may lead to zero or inconsistent values in the original

database, or the data may be tampered with. These wrong data will affect the final results. Therefore, processing before data mining is very necessary.

3.3.3. *Data Integration.* Data integration is the integration of multiple data sources extracted. This article includes basic student information and student physical health evaluation data. It is necessary to use database technology to integrate various database files collected into student physical test data after the above data integration, as shown in Table 1.

TABLE 1: Student basic information table.

Student ID	Height	Weight	Vital capacity	Step test	Grip	Sitting forward bending	Standing long jump
3130408	182.5	78	4330	56	52.3	Null	2.8
3130409	176.2	69	3256	48	Null	32.5	2.5
3130410	156.2	52	4236	89	Null	7.6	1.8
3130411	165.3	51	3695	78	Null	7.9	1.7
3130412	172.6	56	3526	56	54.2	Null	1.9
3130413	181.2	74	2569	89	53.6	7.99	2.1
3130414	168.2	62	3658	68	Null	12.56	1.56
3130415	159.2	60	4562	69	Null	38	1.73
3130416	162	53	3698	71	58.7	Null	Null
3130517	171	58	3879	76	Null	26	1.72
...	...	...	...	...	...	...	...

3.3.4. *Data Cleaning.* In the six items of height forward bending, the total score is calculated according to the National Student Health Standards, the items were divided into four types of “excellent, good, passing, and failing,” further measures were taken to deal with the missing record sheets, and finally, a complete record sheet was obtained.

3.3.5. *Data Conversion.* The side scores of boys and girls are shown in Tables 2 and 3.

Here, we need to perform discretization operations on all continuous features, that is, the classification of the physical status assessment data of pupils. Discretize the total score of the student’s academic year: 90 or greater is excellent, 80 to 89.9 is good, 60 to 79.9 is pass, and less than or equal to 59.9 is fail. After calculating the total score after the data conversion ratio, excellent is counted as 1, pass is counted as 0, good is counted as 3, and fail is counted as -1, and the processed data table is obtained, as shown in Table 4.

From the data in Table 4, it can be obtained that the pass rate of the discretization test of pupils under normal weight is the highest. Among them, only the vital capacity is unqualified, and the pass rate of the discretization test of pupils under overweight is the lowest. Therefore, the test scores of primary school students are closely related to their physical fitness.

## 4. Evaluation and Analysis of the Application Effect of Data Mining Technology in the Public Welfare Sports System

4.1. *Performance Analysis of Data Mining Algorithms.* The experimental data include the number of various files (for more accurate experimental results, for different types of files of similar size, uniformly intercept the same number of data for experiments) and the time consumed by the two algorithms in parallel to process data. There are three data nodes in a fixed cluster in a parallel environment here. In the experiment, various types of data are taken, and serial and parallel experiments are carried out. To compare and analyze these data, Figure 5 shows the time required to run the serial algorithm and the parallel algorithm on the stored identification dataset, Figure 6 shows the time required to run the serial algorithm and the concurrent algorithm on the identification dataset, and Figure 7 shows the time required

to run the serial algorithm and the parallel algorithm on the physical state recognition dataset. From the graph generated by the table, it can be seen that, with a fixed cluster size, when the data size is small, the serial algorithm is more efficient. The advantage of the serial algorithm for processing small-scale data is that there is no data preprocessing operation, there is no data map, and phase delay is reduced. When the algorithm execution time is less than or equivalent to the data processing and result merging time, the serial algorithm can be efficient to perform algorithmic processing. When the data scale is large and the algorithm execution time is much longer than the sum of data preprocessing, distribution to each node, and result merging time, the advantages of parallel algorithms are particularly obvious. At this time, by increasing the number of nodes, the execution efficiency of parallel algorithms can be greatly improved, as shown in Figures 5–7.

In addition, the running time of the algorithm is not only related to the data scale and algorithm parameters but also related to the memory, processing system, and node load service operation of each node in the cluster. Quantitative analysis will not be performed here.

4.2. *Analysis of the Results.* The manual management of student information and the use of data mining technology to organize student information were compared, and the data and information of the students in the six items of height and sitting forward bending were compared, respectively; we will find that it is more efficient to use data mining technology to organize student information. At the same time, the excellent rate, pass rate, good rate, and fail rate of these six projects are shown in Table 5.

The comparison between manual management of student information and the use of data mining technology to organize the accuracy of student information is shown in Figure 8.

4.3. *Analysis of Application Effects.* A large amount of data is generated during nonprofit sports teaching, providing meaningful information for students, teachers, parents, and system developers. This research field belongs to an aspect of educational data mining (EDM). These interesting patterns are effective, novel, and potentially useful for testing data



TABLE 2: Student physical test scores (boys).

Height and weight	Vital capacity	Step test	Grip	Standing long jump	Total score
55	11	65	62	100	72
55	35	75	56	103	82
110	56	68	46	98	87
110	45	81	63	97	76
60	62	92	68	78	76
60	35	87	61	82	84
50	46	56	23	89	82
50	26	61	68	100	83
100	34	59	0	103	79
100	59	72	62	78	80

TABLE 3: Student physical test scores (girls).

Height and weight	Vital capacity	Step test	Grip	Sitting forward bending	Total score
55	62	45	33	92	70
55	67	42	36	65	69
50	12	36	34	87	72
50	15	42	46	69	75
100	19	36	62	63	68
100	56	48	59	79	71
50	45	36	58	81	62
50	56	34	56	75	70
110	35	32	49	80	60
110	23	31	48	69	59

TABLE 4: Physical test results after discretization.

Height and weight	Vital capacity	Step test	Grip	Sitting forward bending	Standing long jump	Vital capacity
Obesity	1	3	0	-1	-1	0
Malnutrition	3	1	0	-1	1	3
Malnutrition	3	1	3	0	1	0
Overweight	-1	-1	3	1		
Normal weight	1	0	3	3	0	0
Overweight	3	1	-1	-1	1	3
Normal weight	-1	-1	0	1	-1	1
Overweight	0	1	-1	0	1	-1
Obesity	3	-1	-1	3	3	1
Normal weight	-1	0	0	1	0	3

and easy to be understood. Data mining technology is the product of multiple disciplines. Through inductive reasoning on a large amount of data, some potential and valuable information and knowledge can be discovered and used as decision support, which can help decision makers find laws and predict future trends. Figures 9 and 10 show how the decision tree reflects the physical health of a student. The decision trees generated by male physical test data are shown in Figure 9.

The most important factor in male physical fitness test is vital capacity, while for students with weak vital capacity, the biggest factor that affects this is height and weight. Vital capacity can usually explain the number of students' extracurricular exercises, and the performance of height and weight can also explain the current physical condition of students. For students with good vital capacity, their vital capacity scores are better; this

is due to normal exercise, and other factors may have better results. For students with poor vital capacity scores, students who are often overweight or obese must pay special attention to physical exercise. The decision trees generated by the female physical test data are shown in Figure 10.

For girls, the most important factor in the physical condition test is the step test, and for students who excel in the step test, the most important factor is vital capacity. The test procedure describes the physical condition of the student, and the vital capacity usually represents the amount of extracurricular exercise. For students who have not yet reached the optimal physical condition, they can use their physical condition to measure their physical condition. For students with better grades, vitality should be used to further reflect the physical health of the students, which is related to normal physical exercise.

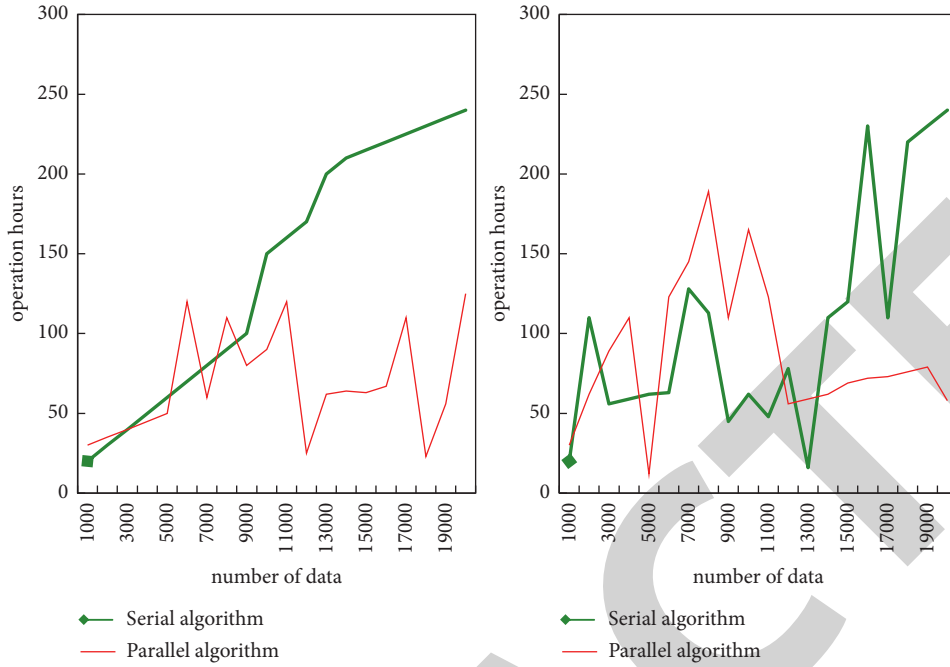


FIGURE 5: Running time of the algorithm for storing identification data.

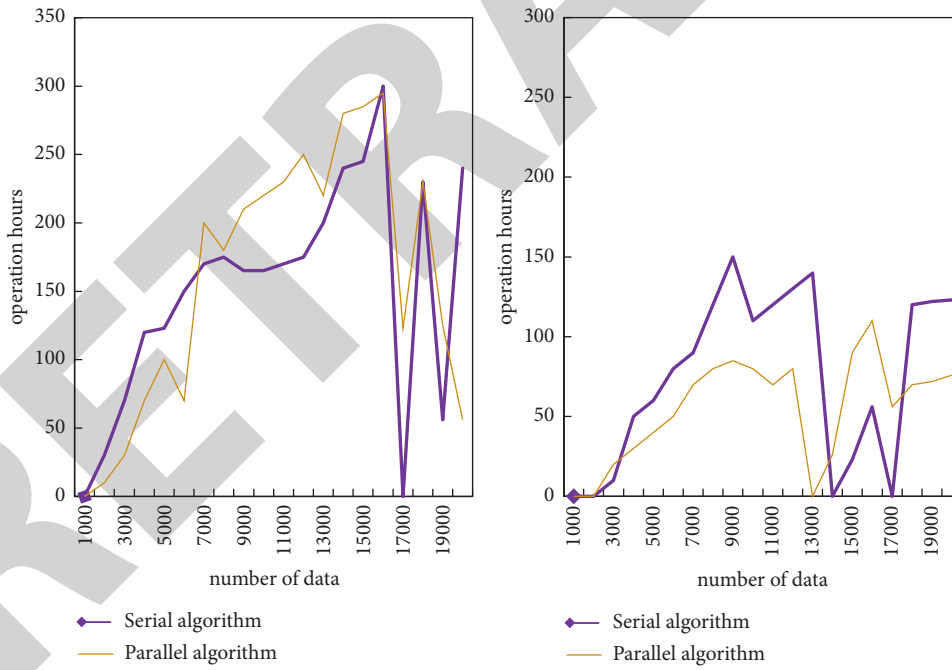


FIGURE 6: Running time of the algorithm for identifying data.

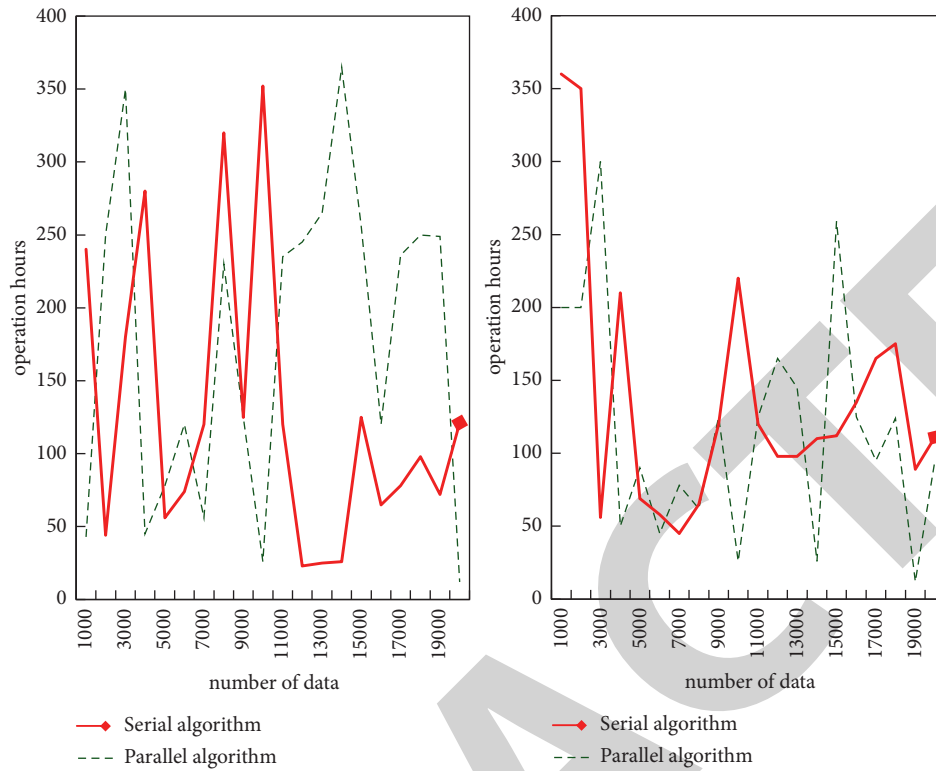


FIGURE 7: Running time of the data algorithm for physical condition recognition.

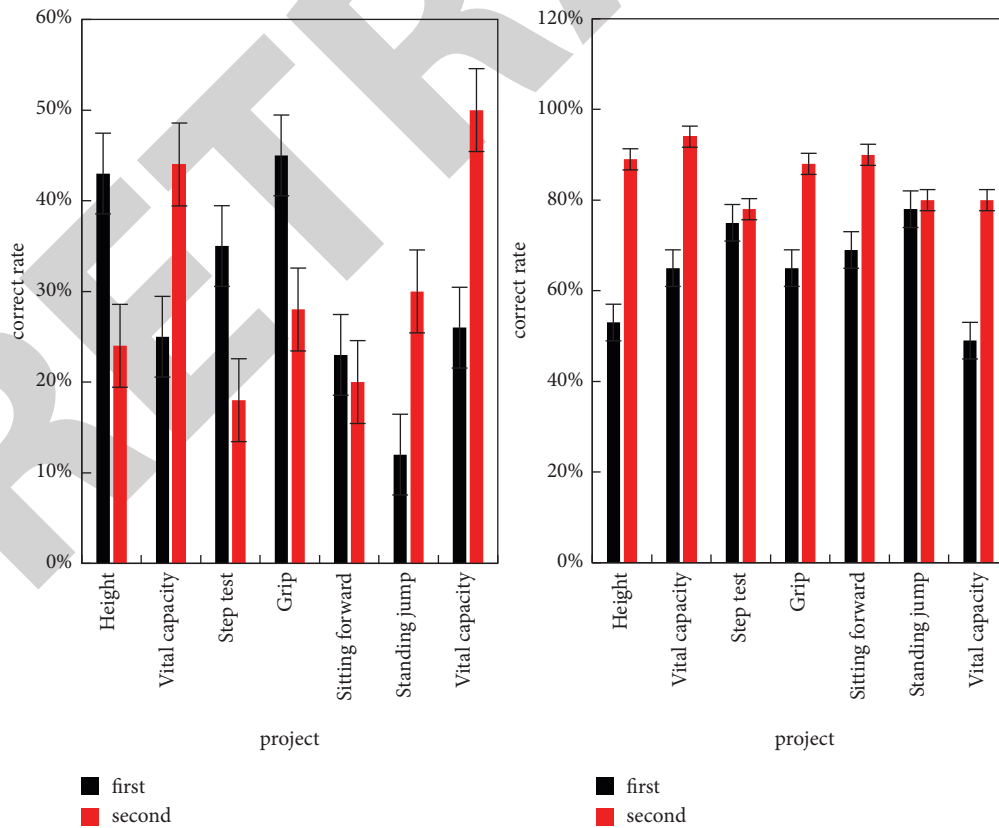


FIGURE 8: Comparison of the accuracy of student information.

TABLE 5: Student body side information result table.

Height and weight	Vital capacity (%)	Step test (%)	Grip (%)	Sitting forward bending (%)	Standing long jump (%)	Vital capacity (%)
Fail rate	23	32	8	12	9	12
Pass rate	58	65	78	78	85	65
Good rate	83	79	85	72	87	89
Excellent rate	19	9	12	13	23	29

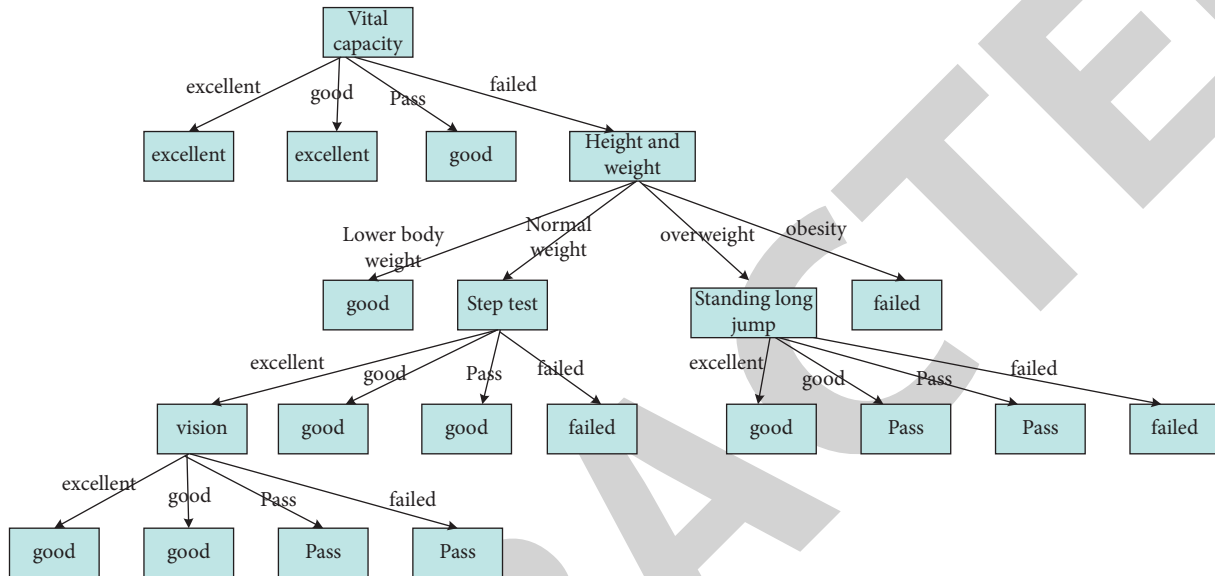


FIGURE 9: Decision tree generated from male physical test data.

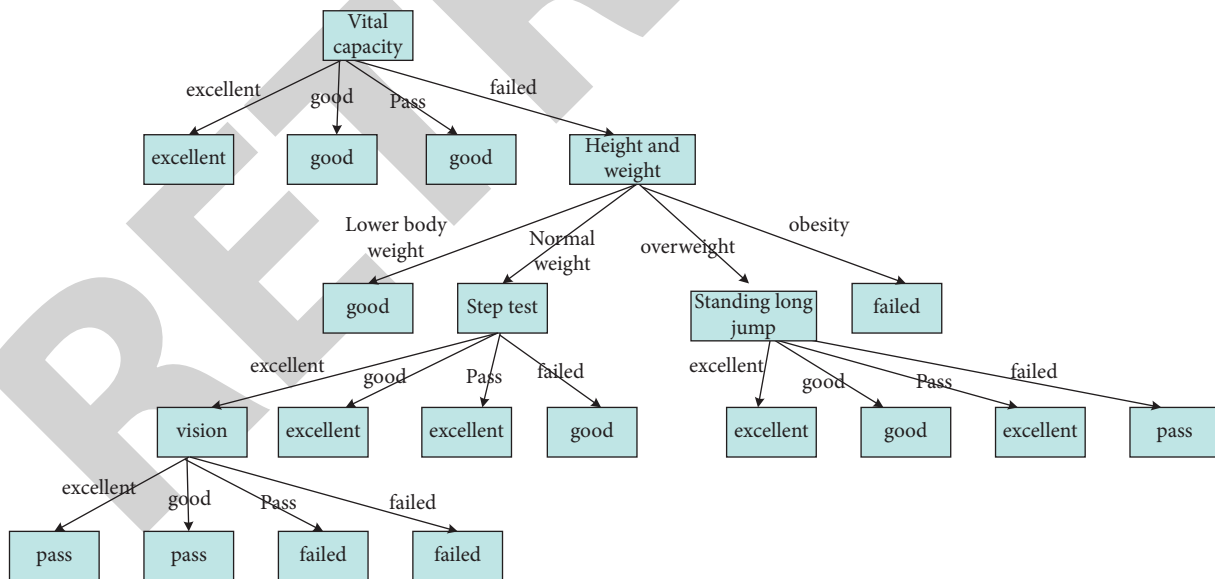


FIGURE 10: Decision tree generated from female physical test data.

## 5. Conclusions

This article first puts forward the research background and research significance; secondly, it introduces the related theories and technologies of data mining and related theories and existing problems in nonprofit sports teaching; then, using the BP neural network, decision tree algorithm, clustering algorithm, and other data mining methods, the process of analyzing the relevant teaching data of the physical education platform is discussed, and the application and analysis of data mining technology in nonprofit sports teaching are emphatically discussed. As the main component of the platform, data resources are analyzed through a large amount of data in the platform to obtain resources that are helpful for different students to improve their physical fitness and constantly improve the platform's functions. Nowadays, the characteristics of computer technology and Internet technology are very obvious, not only is it convenient to use but also can greatly improve work efficiency. Explosion is that the data of all walks of life are growing explosively, and these massive amounts of data are stored in such a deep sleep. However, if you want these data to no longer sleep. Let it create corresponding value for people and serve people's life and production we must redevelop a technology that can solve such problems [30, 31]. However, there are still many shortcomings in the research: first, the data content is not rich enough, which leads to the insufficient breadth of data analysis; second, the data analysis perspective of physical education teaching is not broad enough; the third is to continue to study mining algorithms and design an effective nonprofit sports teaching model suitable for most colleges and universities; as long as the relevant data of sports events are imported, data analysis can be carried out quickly, and effective measures to improve the physical fitness of students can be proposed.

## Data Availability

The data underlying the results presented in this study are available within the article.

## Disclosure

The authors confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

All authors have seen the manuscript and approved to submit to the journal.

## Acknowledgments

This work was supported by the Humanities and Social Sciences (Pedagogy) Youth Fund Project from the Ministry of Education "Research on the Influence of Public Welfare Sports on College Students' Sports Interest and Behaviors

and the Construction of Long-Term Mechanism" (17YJC880020) and NingboTech University of School Education Reform Project (NITJG-201905).

## References

- [1] C. Helma, T. Cramer, S. Kramer, and L. de Raedt, "Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds," *Journal of Chemical Information and Computer*, vol. 35, no. 4, pp. 1402–1411, 2018.
- [2] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017.
- [3] J. Zhang and Y. Li, "Research on the application of data mining technology in the employment and entrepreneurship guidance of university students," *Clinica Chimica Acta*, vol. 42, no. 3, pp. 1059–1063, 2017.
- [4] B. Liang, "Application of network data mining technology in analysing the political education reform and development for higher vocational students," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 18, pp. 114–120, 2017.
- [5] J. Li, H. Chen, and F. Ren, "Application of data mining clustering technology in college and university political education optimization," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 8, pp. 495–502, 2017.
- [6] Z. Tao, W. Jing, and W. Xiaojiang, "Application of data mining technology in university information management," *Agro Food Industry Hi Tech*, vol. 28, no. 1, pp. 1734–1737, 2017.
- [7] M. S. Hemantkumar and A. Adholiya, "A survey on educational data mining using association rule mining' in the," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 3, no. 8, pp. 451–453, 2018.
- [8] L. Lu and J. Zhou, "Research on mining of applied mathematics educational resources based on edge computing and data stream classification," *Mobile Information Systems*, vol. 2021, no. 7, Article ID 5542718, 8 pages, 2021.
- [9] P. Kaur, M. Singh, and G. S. Josan, "Comparative analysis of rank aggregation techniques for metasearch using genetic algorithm," *Education and Information Technologies*, vol. 22, no. 3, pp. 1–19, 2017.
- [10] K. S. Kostyukova, "Digital transformation policy in Japan: the case of artificial intelligence," *MIR (Modernization. Innovation. Research)*, vol. 10, no. 4, pp. 516–529, 2019.
- [11] S. Kozlovskiy, I. Petrunenko, V. Baidala, V. Myronchuk, and T. Kulinich, "Assessment of public welfare in Ukraine in the context of the COVID-19 pandemic and economy digitalization," *Problems and Perspectives in Management*, vol. 19, no. 1, pp. 416–431, 2021.
- [12] C. Li, Y. Shi, P. Gao, Y. Shen, C. Ma, and D. Shi, "Diagnostic model of low visibility events based on C4.5 algorithm," *Open Physics*, vol. 18, no. 1, pp. 33–39, 2020.
- [13] A. M. Khan, D. W. Yun, M. A. Ali et al., "Passivity based adaptive control for upper extremity assist exoskeleton," *International Journal of Control, Automation and Systems*, vol. 14, no. 1, pp. 291–300, 2016.
- [14] T. Gruene-Yanoff and R. Hertwig, "Nudge versus boost: how coherent are policy and theory?" *Minds & Machines*, vol. 26, no. 1-2, pp. 149–183, 2016.

- [15] B. Yu, L. Zhang, F. Guan, Z. Peng, and B. Yao, "Equity based congestion pricing: considering the constraint of alternative path," *Operational Research*, vol. 17, no. 1, pp. 313–337, 2016.
- [16] T. Le, T. C. Son, and E. Pontelli, "Multi-context systems with preferences," *Fundamenta Informaticae*, vol. 158, no. 1-3, pp. 171–216, 2018.
- [17] M. Ryan, "Ethics of using AI and big data in agriculture: the case of a large agriculture multinational," *The Orbit Journal*, vol. 2, no. 2, pp. 1–27, 2019.
- [18] C. Zambrano and Y. Olaya, "An agent-based simulation approach to congestion management for the Colombian electricity market," *Annals of Operations Research*, vol. 258, no. 2, pp. 217–236, 2017.
- [19] M. M. Ozciloglu, M. F. Akay, and D. Heil, "On the application of artificial intelligence and feature selection in sports science education and research: a case study," *New Trends and Issues Proceedings on Humanities and Social Sciences*, vol. 3, no. 3, pp. 256–262, 2017.
- [20] J. L. D. Vera, M. D. Andrada, A. Q. Bello, and M. G. de Vera, "Teachers' competencies in educational technology integration on instructional methodologies in the new normal," *International Journal of Online Pedagogy and Course Design*, vol. 1, no. 1, pp. 61–80, 2021.
- [21] Z. Lu, "Research on the application of computer data mining technology in the era of big data," *Journal of Physics: Conference Series*, vol. 1744, no. 4, Article ID 042118, 2021.
- [22] W. Yan, J. Tan, H. Zhan, and H. Wang, "Research on the method of fault diagnosis based on multiple classifiers fusion," *International Journal of Hybrid Information Technology*, vol. 9, no. 2, pp. 195–202, 2016.
- [23] L. N. M. Bezerra and M. Terra da Silva, "Application of EDM to understand the online students' behavioral pattern," *Journal of Information Technology Research*, vol. 12, no. 3, pp. 154–168, 2019.
- [24] C. Alfaro, J. Cano-Montero, J. Gómez, J. M. Moguerza, and F. Ortega, "A multi-stage method for content classification and opinion mining on weblog comments," *Annals of Operations Research*, vol. 236, no. 1, pp. 197–213, 2016.
- [25] Y. Huang, P. Swamidass, and D. A. Raju, "The nature of innovation in emerging industries in China: an exploratory study," *The Journal of Technology Transfer*, vol. 41, no. 3, pp. 451–468, 2016.
- [26] D. Gong, L. Feng, L. Xiao-Ting et al., "The application of S-transform spectrum decomposition technique in extraction of weak seismic signals," *Chinese Journal of Geophysics*, vol. 59, no. 1, pp. 43–53, 2016.
- [27] A. Kamaz, K. Yildiz, and A. Buldu, "Aratrma makalesi/research article an application on technology addiction with C4.5 classification algorithm," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 9, no. 4, pp. 1756–1765, 2020.
- [28] X. Zhang, "Evaluating the quality of internet-based education in colleges using the regression algorithm," *Mobile Information Systems*, vol. 2021, no. 1, Article ID 7055114, 9 pages, 2021.
- [29] J. Tanha, M. Van Someren, and H. Afsarmanesh, "Semi-supervised self-training for decision tree classifiers," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 1, pp. 355–370, 2017.
- [30] A. Linden and P. R. Yarnold, "Modeling time-to-event (survival) data using classification tree analysis," *Journal of Evaluation in Clinical Practice*, vol. 23, no. 6, pp. 1299–1308, 2017.
- [31] A. I. T. Tulloch, N. Auerbach, S. Avery-Gomm et al., "A decision tree for assessing the risks and benefits of publishing biodiversity data," *Nature Ecology & Evolution*, vol. 2, no. 8, pp. 1209–1217, 2018.