With the popularization of intelligent devices and the rapid development of information technology, many valuable data become easier to obtain. After obtaining massive data, the problem of how to make more efficient and scientific use of these data is in front of us. Data mining is a deeper processing of data. It is aimed at mining the internal associations hidden in complex data and making the data give full play to its maximum value. At present, college physical training is also in urgent need of more scientific and modern upgrading and reform. Data mining is introduced into college physical training so that physical training data is no longer just a simple collection, query, and storage. This paper uses and introduces the basic content of Apriori algorithm. The actual college sports training data are collected for preprocessing, and the discrete dynamic model is established by using Apriori algorithm. Using this model, the actual data are deeply analyzed. Using the strong correlation subintegration results, the physical training of college students is optimized and reformed. The scores of college students’ physical quality are improved, and the goal of sustainable development of college physical training is achieved. Based on the data of 126 students in four classes of a major in 2020, this paper mines the frequent subsets with strong correlation in the data and finds out the relationship between the correlation and trust between subsets. In the 70-day experiment, the scores of college students’ comprehensive physical quality have been significantly improved, which proves that the discrete dynamic model established by Apriori algorithm data mining has a significant role in college physical training. The questionnaire survey results show that 83% of the college students are very satisfied with the discrete dynamic model-assisted physical training, and 16% of the college students are basically satisfied with the system. Apriori algorithm has significant advantages in dealing with big data. Single Apriori algorithm also has some defects. The efficiency of data mining, data security, and the accuracy of model analysis results need to be tested in practice.
also needed as a means to give full play to the value of big data. Nowadays, various popular and general algorithms include decision tree algorithm that can process incomplete data. There is a naive Bayesian algorithm based on the principle of probability theory. There are vector machine algorithms that can map low-dimensional data to high-dimensional data. K nearest neighbor algorithm with K neighbors represents the nearest sample. There is an AdaBoost adaptive lifting algorithm that can combine multiple weak classifiers into a strong classifier. There is a widely used Apriori mining association algorithm to reveal the association between samples. There are also EM clustering algorithms to establish core reclassification. There is also the PageRank algorithm which originated from the influence calculation of the paper [2–9]. Using algorithm technology to mine real data is the ultimate goal of big data processing. Only in this way can we reflect the ultimate value of big data.

Data mining has been applied in the government and enterprises for a long time and has also achieved quite good results and achievements so that the government can better serve the people, and enterprises have made progress in product promotion, user portrait, and advertising [10]. Another data-intensive place, where data mining has just started and has not yet used algorithm technology on a large scale, is the major universities in China. The storage and use of some public data in colleges and universities are still inefficient, resulting in the waste of teaching resources and management in the last century. In addition, some colleges and universities have mined the data of students’ cultural course scores, classroom performance, and dormitory conditions, so as to improve the quality of teaching and student management. However, sports training, which has a large amount of data to be mined, has not received enough attention [11]. Therefore, this study is aimed at selecting the appropriate algorithm and conducting discrete dynamic modeling and analysis on the real college physical training data through the selected Apriori algorithm so that physical education teachers and relevant educators can not only see the students’ achievements in a certain time or speed but also carry out more means from more aspects to improve the students’ all-round physical education. Finally, the discrete dynamic model established by data mining becomes a useful tool for improving sports performance [12].

The significance of using data mining algorithm to establish discrete dynamic model for college physical training lies in the following three aspects. The first aspect is that compared with manual data processing, the efficiency of discrete dynamic model will be greatly improved, the subjectivity of manual data processing is avoided, and the authenticity and objectivity of data processing structure are guaranteed [13]. On the other hand, data mining can comprehensively consider the relevance of various indicators. For example, it can organically correlate and analyze the indicators such as college students’ basic physical fitness, test results, and after-school training, so as to minimize the one-sidedness of the results [14]. The last aspect is that the discrete dynamic model can be tracked and calculated for a long time, from one week to the whole academic year, and can be analyzed regularly, which provides a good basis for the evaluation of teaching quality for educational decision-makers [15].

2. Selection and Introduction of Apriori Algorithm

Aiming at the above concepts and problems, this paper establishes a discrete dynamic model based on Apriori algorithm, a widely used data mining algorithm. By integrating the model into the teaching information management system, collecting students’ basic information and physical training, this paper analyzes the college physical training data. Using Apriori algorithm, the relationship rules between different items of college physical education learning are obtained, and the weak links of students are calculated quickly and accurately. Teachers can arrange teaching and training tasks pertinently. The final results, in the use of a period of time, show superior characteristics, making college physical training achieve the strategic goal of sustainable development.

Nowadays, the algorithm is not only more than ten kinds mentioned above but also evolving and updating in order to meet the new requirements. Apriori algorithm has been widely used in the frontline of data processing since its invention. This study chooses it for the following three reasons: first, college physical training data is a sparse data set, which is highly consistent with Apriori algorithm. Secondly, the principle of Apriori algorithm is simple and easy to realize the research requirements. Finally, the college students’ physical fitness data, training intensity data, test score data, and after-school exercise data collected in college physical training have strong correlation and complex internal relationship, so it is suitable to use Apriori algorithm to establish discrete dynamic modeling [16].

In data mining, Apriori algorithm is a method to find the association rules between data samples. It can find the frequent data sets hidden in big data. The most classic case of Apriori algorithm in the research is to find the internal relationship between diapers and beer, put them together on supermarket shelves, and improve the sales of two commodities at the same time, which is the embodiment of Apriori algorithm in favor of decision-making. The iteration of Apriori algorithm is carried out layer by layer. First find out frequent data set 1, and then find frequent data set 2 through data set 1, and so on. Apriori algorithm uses the implication expression from X to y to reflect the association rules of two disjoint items sets X and Y. It uses the support rate and confidence rate to measure the association strength between X and y.

\[ S(X, Y) = P(X, Y) = \frac{\text{num}(xy)}{\text{num}(all \ samples)}, \]

\[ C(XY) = P(x|Y) = \frac{P(xy)}{P(y)}. \]

S and C in formulas (1) and (2) represent support rating and confidence rating, respectively, and num represents the total set.

2.1. Data Collection and Preprocessing of Physical Training in Colleges and Universities. In the traditional teaching data collection of colleges and universities, the main component
is the achievement statistics of college students. This traditional collection only requires teachers to input the data into the form of office software and then add, subtract, and rank through the simple Excel self-contained formula. The data of physical training in colleges and universities should be more three-dimensional and multifaceted, with significant information characteristics. In the new era of data collection with big data support, we pay more attention to the cooperation with a variety of sensors and teaching management platforms and make use of existing resources to achieve the strategic goal of sustainable development. Firstly, the experimental data determined in this study mainly include the basic information of college students, the training in physical education, the training after physical education, and the results of midterm and final physical examination and sports competition. Among them, the student name, student number, gender, major name, and compulsory elective courses in the basic information of college students will be entered or assigned at the time of enrollment, which can be obtained through the teaching management platform of colleges and universities [17]. The sports preference, hobbies, and personality of college students can be obtained by questionnaire. The situation in physical education class is recorded by physical education teachers and physical education committee members in class and recorded in the database of teaching management platform after class [18]. The training situation after physical education class is collected by college students by entering the "college physical education independent monitoring training data system" (hereinafter referred to as the automatic control data system). The results of sports competitions are entered into the teaching management database by the teachers who lead the team. If individuals participate in the competition, they will be reported to the teachers for entry.

It can be seen that the original data collection of college physical training mainly depends on the teaching information management platform and automatic control data system. Compared with the traditional excel table statistics, the teaching information management system is more intelligent [19]. It adopts CS hybrid architecture, integrates modules of data storage, analysis, and output in the system, and can process a large number of relatively complex student data. Similarly, the architecture of automatic control data system is also completed by CS hybrid structure, so the two can be organically combined to be competent for data acquisition and daily use of teachers and students. The common advantage of the two systems is that the system administrator can directly operate the server without transfer, issue instructions, and download

the required information. The structural layout of teaching information management platform and automatic control data system is shown in Figure 1.

According to the teaching information management platform and automatic control data system shown in the above structure diagram, it can be seen that the application server and data server can interact with each other by themselves. The application data server can be directly used by the platform administrator or filtered by the firewall by teachers and students through the automatic control data system.

Apriori algorithm is different from the old-fashioned data processing methods and cannot carry out violent search. Therefore, it is necessary to prune the collected college physical training data. This operation is mainly based on two theorems of Apriori algorithm. The first theorem of Apriori algorithm: if a set is a frequent itemset, all its subsets are frequent itemsets. The second theorem of Apriori algorithm: if a set is not a frequent itemset, all its supersets are not frequent itemsets. As shown in Figure 2, first continuously subdivide the original data into several subsets and finally merge the subsets into a collection of effective factors. Then, find the subset AB that is not a frequent itemset. According to the first and second theorems of Apriori algorithm, we can know that the sets containing subset AB are not frequent itemsets, that is, the subsets within the dotted line range in the graph. Finally, the pruning operation is completed by removing this part. Since the scope of this study is college sports training data, pruning usually includes the information of transfer and dropout students, invalid classroom after-school training data, invalid midterm and final examination scores, and invalid competition scores.

After pruning the original data, because the data obtained by the teaching information management application server is heterogeneous, it also needs to be normalized. First, use matrix formula (3) to preliminarily classify and arrange the data.

$$D_T = \begin{bmatrix}
    d(x_1, x_1) & \cdots & d(x_1, x_T) \\
    \vdots & \ddots & \vdots \\
    d(x_T, x_1) & \cdots & d(x_T, x_T)
\end{bmatrix}.$$  \quad (3)

$T$ in the formula represents the total quantity of all data, and then, use formula (4) to obtain the average value of the data.

$$R = \frac{\sum_{i=1}^{T} R_i}{T} \quad (i = 1, 2, \cdots, T).$$  \quad (4)
\( \bar{R} \) in the above formula represents the average value of \( R_i \) with a total number of \( T \). Finally, the data is normalized by weighting the error according to the classification. The following formula needs to be used.

\[
R = \bar{R} + d_{\min}(x_i, x_j).
\]  

(5)

After pruning and normalization, the data structure is unified, which is easier to mine and model later. Apriori algorithm can also use machine learning method to compensate the missing necessary data. The method of supplementing data is mainly divided into two steps. The first step is to compare the missing students’ data with all other students and select some key factors, such as students’ height and weight data and physical training volume. Find no less than 50 samples of similar data. The second step is to predict the 50 samples obtained, so as to scientifically calculate the missing data. This study uses the data of a major in 2020 in a university to complete the hypothetical deficiency. Firstly, hide the key information of three students, and compare the inferred data with the real data using the above method. The results are shown in the figure.

The accuracy rate is obtained by dividing the speculated data made by the above method by the real data. From Figure 3, it can be seen that the speculated accuracy rate of the basic personal information and physical training data of the selected three students A, B, and C is more than 95%. The research results show that the Apriori algorithm is scientific and reliable to compensate for the missing data.

2.2. Apriori Algorithm for Discrete Dynamic Modeling and Example Analysis

2.2.1. Establishment of Discrete Dynamic Model. In order to master the strengthening degree of diversified college students’ physical training on different students’ physical functions, Apriori algorithm is used for data mining and discrete dynamic modeling of college students’ physical training data in a college, in order to achieve the purpose of sustainable development [20]. In the construction of discrete dynamic model, considering different training items in physical education class and different competition items in sports competition, the prediction vector of data positioning state is

\[
\alpha = (a_1, a_2, \cdots, a_n) \neq 0.
\]  

(6)

According to the strategic goal of college students’ sports training plan, the diversified college students’ sports training data is transmitted to the application server. When
scheduling the data, the characteristic vector is used to refer to
\[ x^{(k)} = \begin{bmatrix} x_1^{(k)} & x_2^{(k)} & \cdots & x_{N_k-1}^{(k)} \end{bmatrix}^T, \]
\[ s^{(k)} = \begin{bmatrix} s_1^{(k)} & s_2^{(k)} & \cdots & s_{N_k}^{(k)} \end{bmatrix}^T, \]
\[ y^{(k)} = \begin{bmatrix} y_1^{(k)} & y_2^{(k)} & \cdots & y_{N_k}^{(k)} \end{bmatrix}^T. \]

In the formula, \( x^{(k)} \) and \( y^{(k)} \) represent the linear horizontal and vertical two-way input of the system, and \( s^{(k)} \) represents the reversible and invariant output.

The results of sports competition under comprehensive exercise are regarded as the time-frequency characteristics, which reflect the essential characteristics of college physical training. The collection of all data is \( N \) discrete dynamic distribution points \( P = \{p_1, p_2, \ldots, p_N\} \). The time mean and frequency mean of the gain index of college students participating in the physical training of this experiment can be calculated as the following formulas:
\[ t_m = \frac{1}{E} \int_{-\infty}^{\infty} t(x(t))^2 \, dt, \]
\[ v_m = \frac{1}{E} \int_{-\infty}^{\infty} v[X(x)]^2 \, dv. \]

After the above series of processing, we get the college physical training index system using linear or approximate linear discrete dynamic model. Fractal tree structure of college students’ physical training information storage in order to analyze the discrete dynamic model of the promotion relationship between college physical education courses and the physical gain index of college students, based on the examination of the whole academic year and the data collection and analysis uploaded by teachers and students, master and regularly adjust the physical improvement degree of physical training plans for different students. Based on this model, the principal component analysis method is used to calculate the variance matrix \( C \):
\[ C = \frac{1}{N} [X - \bar{X}] [X - \bar{X}]^T. \]

Establish a linear discrete dynamic system to fit a variety of college physical training factors to realize the fitting of diversified data in the discrete dynamic model. The model can be expressed by the following formulas:
\[ R_{\beta}X = U \left\{ E \in U \mid c(E, X) \leq \beta \right\}, \]
\[ R_{\beta}X = U \left\{ E \in U \mid c(E, X) \leq 1 - \beta \right\}, \]
\[ bnr_{\beta}(X) = R_{\beta}X - R_{\beta}X_1. \]

It is necessary to merge frequent subsets into the parent set with universal influence and select the number of main sets according to the influence degree of variance accumulation. Only when the variance accumulation reaches a certain contribution, the corresponding subsets can be merged into the universal parent set. Through the above discrete dynamic fractal design, this experiment realizes the establishment of the model of college physical training data mining.

2.2.2. Case Analysis of Physical Training Data in a University. In order to verify the analysis effect of discrete dynamic modeling established by Apriori algorithm on college students’ physical training, a total of 126 students in four classes...
of a major in 2020 in a university are taken as objects, and the discrete dynamic model is used to test and analyze the physical training effect of these students. The model uses the full score system to score the frequent subsets in the data. For example, students’ endurance score of 10 is particularly excellent, 8 is good, 6 is medium, and 4 is poor. The scoring system can quantify the ability of students’ specific data with numbers. Data mining of 126 students using discrete dynamic model can find. The interests and hobbies of college students are strongly related to the examination results of some items. In this study, support is used to express the correlation degree of two frequent subsets, and trust is also used to express the possibility of high scores. The curve of relevance and trust can be made through the model, as shown in the figure.

As can be seen from Figure 4, the higher the correlation degree of the two frequent subsets, the higher the trust degree of the score subset, that is, the greater the probability of getting high scores. When the believability degree is lower than 64.3%, the trust degree increases rapidly. When it is higher than 64.3%, the growth rate of trust degree slows down, but it still maintains a positive correlation. Through detailed analysis, we can know that when a student’s hobby is jogging, his endurance test score subset is strongly related to this hobby subset. When the correlation between hobbies and endurance test is more than 80%, the trust can reach more than 70%, and then, the student’s endurance test score subset is usually more than 7 points. It can also be pushed back. When another student obtains a medal or reaches 9 points in table tennis, the correlation between his physical reaction speed and table tennis performance will be greater than 90%. Through the discrete dynamic model, the subsets with high correlation degree in the data are mined out. The physical education teachers of each class can increase the amount of training for the weak items shared by the whole class and can also reasonably allocate to participate in sports competitions according to the advantageous items of a student. On the basis of mining the subset with high correlation degree by using the discrete dynamic model, this study cooperates with teachers and students to carry out phased courses, makes scientific adjustment according to the data fed back by the model, and obtains the changes of four grades after 70 days of recording and processing, as shown in the figure.

As can be seen in Figure 5, after using the discrete dynamic model to optimize the college physical education curriculum training, within 70 days, the endurance, physical reaction speed, flexibility, and core strength of college students have been significantly improved, which shows that the model obtained by data mining using Apriori algorithm in this experiment has made an expected contribution to the sustainable development of college physical training. After completing the 70-day experiment, this study conducted a satisfaction survey on the college students participating in the experiment by issuing a questionnaire. As shown in the Figure 6, 83% of the college students were very satisfied with the discrete dynamic model-assisted physical training, and 16% were basically satisfied with the system.

The data mining efficiency of traditional PageRank algorithm and Apriori algorithm in discrete dynamic model is tested. It can be seen that the number of samples, that is, the number of college students using this discrete dynamic model, directly affects the efficiency of the two algorithms, and the results are shown in Figure 7.

As can be seen from Figure 7, when the number of samples is 20, 40, and 60, the data mining efficiency of PageRank algorithm is higher than that of Apriori algorithm, but when the number of samples is greater than 80, the efficiency of Apriori algorithm is higher than that of PageRank.
algorithm. This shows that the Apriori algorithm also has higher efficiency for more samples. In the era of increasing data explosion, the Apriori algorithm is more in line with the development needs of the times. By analyzing the reasons, we can know that the pruning and normalization of data by Apriori algorithm effectively reduce the interference of data noise, and it is easier to obtain the direct correlation of frequent subsets in the case of massive samples.

3. Conclusions

(1) The data mining algorithm based on frequent subset association establishes a discrete dynamic model based on the basic information, physical exercise information, and physical fitness information of college students, which provides a driving force for sustainable development. In this study, Apriori algorithm is used to prune and normalize the original data to further establish the discrete dynamic model. The model has been integrated into the college teaching information management platform.

(2) Based on the data of 126 students in four classes of a major in 2020, this paper mines the frequent subsets with strong correlation in the data and finds out the relationship between the correlation and trust between subsets. Using the results of strong correlation integration, the physical exercise of college students is optimized and reformed. In the 70-day experiment, college students’ physical fitness scores were significantly improved. It is proved that the discrete dynamic model established by Apriori algorithm data mining has a significant role in improving college physical training. The questionnaire survey results show that 83% of the college students are very satisfied with the system to assist physical training, and 16% of the college students are basically satisfied with the system.

(3) The mining efficiency of Apriori algorithm and PageRank algorithm is compared. The results show that when the number of samples is less than 60, PageRank algorithm is better than Apriori algorithm. When the number of data samples is greater than 60, Apriori algorithm is significantly better than PageRank algorithm in dealing with a large number of samples. Using Apriori algorithm to build a discrete dynamic model and introducing it into the teaching system can regularly monitor the teaching results, objectively reflect the students’ training results, and then make targeted training adjustments. It has made great contributions to the realization of the strategic goal of the sustainable development of college sports.

(4) Apriori algorithm also has some defects. The efficiency and security of data mining and the accuracy of model analysis results need to be tested in practice.

Data Availability

The figures used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to show sincere thanks to those techniques which have contributed to this research.

References


