International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-19

https://doi.org/10.70917/ijcisim-2025-0264
Article

A Study on Classification and Instructional Adjustment of Students' Physical Fitness Test Data Based on Cluster Analysis

Lin Xu *

Qiongtai Normal University, Haikou, Hainan, 571127, China; a13876623667@126.com

Abstract: Enhancing students' physical fitness level is the current focus of physical education teaching, and how to do a good job of teaching physical fitness training is also a problem that physical education teachers must think about. In this paper, the physical fitness test data of undergraduates of grade 2023-2024 of University A in 2025 is selected as an example, and the physical fitness test data is divided into male and female groups, and the improved K-means method is applied to cluster analysis of the two groups of data, and outlier analysis of the clustering results is carried out. Then the buddy recommendation model based on set-pair theory was proposed to get the degree of recommendation between users, and the buddy recommendation model based on set-pair theory was utilized to design the students' physical test teaching intervention experiment, and the data of various indexes before and after the experiment were statistically analyzed. The results of the study show that there are differences between male and female groups in the physical fitness test, and the percentage of outlier students whose physical fitness data is different from the majority is found to be about 1%, and the buddy recommendation model based on set-pair theory has a better holistic effect and stability than the traditional recommendation algorithms in the recommendation of exercise prescription. The experimental conclusion verifies that the exercise prescription generation and intervention monitoring system has a certain role and scientific validity in promoting students' health through exercise intervention.

Keywords: physical health; K-means clustering; exercise prescription; recommendation system; physical fitness test

1. Introduction

The physical fitness status of students is directly related to their physical and mental health and future development [1]. As an important means of assessing students' physical fitness, physical fitness test can provide valuable information for schools, parents and the society. By analyzing the data of physical fitness test, we can understand the current status of students' physical fitness level, find out the existing problems, and put forward the corresponding improvement measures in order to promote the overall improvement of students' physical fitness [2-5].

An official data shows that in 2024, the national youth physical fitness attainment rate is 67.8%, but the percentage of excellent physical fitness is not more than 30%. This seriously exposes the defects of the “centralized physical training” teaching mode of traditional physical education. This centralized teaching mode ignores the individual differences of students, widening the physical fitness gap between students; and teaching content dispersed, teachers' limited energy, unable to achieve personalized guidance, it is difficult to find out the differences in the level of students' physical fitness; and the physical fitness data of the test is only used as a basic reference for students' physical fitness or not, and does not make full use of these data to dig into the characteristics of the student group and individual differences [6-10]. In the national emphasis, how to make full use of the main position of the physical education classroom, through the physical fitness test data to adjust the mode of physical fitness teaching, and effectively improve students' physical fitness and health, is an important issue facing the majority of sports workers.

Nowadays, data-driven education has gradually entered the general education, the use of big data, machine learning and other intelligent technologies for student data analysis, capturing students' learning, physiological, psychological and other characteristics, responding to students' individual differences, schools and teachers based on the results of these analyses to make teaching adjustments to improve students' subjective initiative and interest, which can help to promote the development of personalized and precise education [11-15]. Among them, cluster analysis, as an unsupervised machine learning algorithm, is a set of statistical analysis techniques to classify research objects into relatively homogeneous groups [16]. It has been widely used in the field of data analysis and education because it does not require pre-labeling of sample categories and only uses the similarity or distance of the data itself to group them [17-18]. Cluster analysis method is based on the students' physical fitness test data divided into student groups with different characteristics, which provides a scientific basis for the adjustment of physical education.

Currently, the analysis of physical fitness test data based on cluster analysis has achieved preliminary results. Burns et al. [19] used hierarchical cluster analysis to classify college students' physical fitness test score data and introduced one-way ANOVA to calculate the index differences of the classified groups.The students' percentage of body fat, peak maximal oxygen uptake, and single maximal bench press performance indexes showed significant differences, which helped teachers to develop the physical training program. Gao et al[20] analyzed college students' physical test scores by combining correlation analysis, cluster analysis, and Apriori algorithm, taking physical test indicators as antecedents and mining the association rules therein, and found that the parameters of students' speed, endurance, and agility significantly affected students' physical fitness, which revealed educational weaknesses. Cepkova et al [21] used cluster analysis and analysis of variance (ANOVA) to distinguish three groups of students with similar characteristics based on students' physical fitness test data and anthropometric data, and the identification of these groups' characteristics can help teachers to implement personalized teaching. Gao et al. [22] introduced a cluster analysis method of self-organized feature mapping network that can automatically categorize students' physical fitness test scores into three grades, which intuitively and clearly demonstrates the key factors influencing the physical fitness test scores. The key influencing factors of students' physical fitness are clearly demonstrated in an intuitive way.

Hart [23] used k-mean clustering to analyze the physical fitness measurement variables of college students, and identified four groups of students' physical fitness characteristics, i.e., anaerobic+healthy, aerobic+healthy, overweight+unhealthy, and normal weight+unhealthy. Tang et al [24] used k-mean clustering to classify students' physical fitness test data, combined with support vector regression model to capture the data features, and established a prediction model of students' physical indexes to assess students' physical fitness level, and further constructed a prediction model of students' physical fitness to assist teachers in adjusting their teaching programs. Zhang and Jiao [25] analyzed the data of college students' physical fitness test based on the k-mean clustering algorithm. The results pointed out that the change trend of physical fitness data showed gender differences, with female students' performance being complex, while male students' performance was relatively smooth. Ji et al [26] analyzed the physical fitness data of college students by principal component analysis and improved Canopy-K mean algorithm to group the students' physical fitness test data and draw a physical fitness portrait, pointing out that the students' physical fitness differed among gender, age, and majors, with the Department of Agricultural Sciences>Department of Science and Technology>Department of Humanities and Social Sciences. Feak et al[27] analyzed the relationship between physical fitness of physical education majors and different indicators of physical development by using k-mean clustering method, and the results of the analysis showed that female students have better coordination and speed performance, while male students have better strength and endurance performance, and compared to height, excessive obesity is the negative factor affecting physical health. Guo et al[28] combined the improved k-mean algorithm and machine learning model to construct a physical health portrait based on students' physical fitness data, and based on the characteristics of the portrait, they proposed the corresponding physical fitness intervention strategies for different groups of students.

In addition, Yang [29] conducted a fuzzy cluster analysis of students' physical fitness with the help of a cluster extraction algorithm based on weight, lung capacity, endurance, flexibility, and speed indicators, and performed a hybrid pool feature representation of the indicator features for clustering the physical fitness big data for assessing students' physical fitness. Yan [30] combined a clustering algorithm with association rule mining for students' physical fitness test data, with the former being used to The former is used to classify students' physical fitness characteristics, and the latter is used to mine the associations between physical fitness indicators, so as to effectively assess the physical fitness and health level of highly efficient students, and to support schools to target sports and health management. Tong and Wang [31] reported a fuzzy clustering prediction method of college students' physical fitness to analyze students' physical fitness data, and proposed that based on the results of the analysis, colleges and universities need to improve their campus environments, set up scientific sports facilities, reform physical education, and cultivate students' physical fitness and health. Zeng et al [32] used K-mean cluster analysis to analyze the physical fitness data of nearly 10,000 college students, and designed a number of targeted courses to improve students' physical fitness through the analytical structure, which showed that students' physical fitness and health were improved after a year of teaching adjustments. These studies fully demonstrate the effectiveness of cluster analysis in the analysis of students' physical fitness data, but the data that the studies focus on are missing students' exercise data, recent sleep data, and physiological data, and the underlying K-mean clustering algorithm needs to preset the number of clusters, is sensitive to high-dimensional data, and is prone to lead to classification errors. In addition, most of the studies only analyzed the data and did not use the results of the analysis to adjust the curriculum and teaching models.

In this paper, taking the physical fitness test data of undergraduate students of class 2023-2024 in 2025 as an example, the improved K-means method was chosen to divide the college students' physical fitness test data into two groups of men and women, clustering them into 2-7 classes, and the results were visualized and analyzed. Subsequently, the traditional similarity calculation method is improved, a new set-pair similarity recommendation method is proposed, and a similarity degree-based exercise prescription method is designed. On this basis, experiments were designed to ensure the rationality, safety and effectiveness of the core parameters of exercise prescription. Finally, an 8-week exercise intervention experiment was conducted to collect the index data of the participants before and after the experiment, and the experimental data were statistically analyzed using SPSS software.

2. Classification of student fitness test data based on cluster analysis

2.1. Data sources

In order to obtain more reasonable training results, 5452 data were selected from the 2025 physical fitness test of undergraduates in the classes of 2023-2024 of University A. The age of the test subjects generally ranged from 18 years old to 22 years old, of which 2864 data were measured for male students and 2588 data were measured for female students. The boys' data set contains eight indicators: height, weight, lung capacity, 50m run, standing long jump, sitting forward bending, 1000m run and pull-ups; the girls' data set contains eight indicators: height, weight, lung capacity, 50m run, standing long jump, sitting forward bending, 800m run and sit-ups.

Since the correlation between height and weight and other physical indicators could not be evaluated, the two indicators of height and weight were not chosen as the basis for clustering in this paper. At the same time, in order to avoid the influence of the disparity in the magnitude of each variable value, outliers, and extreme values on the clustering results, the normalization was performed using the method of mapping the matrix to a mean of 0 and a standard deviation of 1 to obtain the values of the dimensionless variables.

2.2. Cluster analysis methods

2.2.1. K-Means Clustering

In this paper, k-Means clustering algorithm is used for algorithmic improvement. K-Means is a divisive clustering algorithm, which is easy to implement and has considerable efficiency [33]. First, K objects are randomly selected from the clustered data as the initial points, also known as the center of mass, corresponding to the K classes; then the distance between each of the remaining objects and the K initial points is calculated, and they are assigned to the initial points with the closest distance to form the initial clustering grouping until the clustering center of mass is no longer changed or the number of iterations meets the requirements. Because the initial clustering is generated on the basis of random selection of the center of gravity, it is impossible to ensure that the resulting K classes is the objective existence of the “natural small class”, so the K-Means algorithm needs to be repeated many times.

The disadvantages of the K-Means algorithm are obvious: firstly, in the K-means algorithm, the value of K must be given by the user, which is arbitrary, and the user can not give the optimal number of clusters. Secondly, in the K-means algorithm, the initial point is either specified by the user or randomly pointed by the system. If the initial point is not well chosen, the algorithm needs to be iterated continuously, and the time overhead will be larger.

2.2.2. Improved K-means clustering algorithm

(1) Improved K-value selection

According to the clustering results of K-Means algorithm, two statistics, total SSE value and total

SSB value, can be calculated.

The total SSE value represents the sum of the sum of squared deviations of all clustered variables with the following formula:

$$SSE = \sum_{i=1}^{k} \sum_{x \in c_i} dist(c_i, x)^2 \tag{1}$$

where c_i is the i th cluster, x is the point in c_i, c_i is the mean of the i th cluster, and $dist$ is the distance between two objects.

The total SSB denotes the sum of squared deviations of the clustering variables between the categories with the following formula:

$$SSB = \sum_{i=1}^{k} m_i dist(c_i, c)^2 \tag{2}$$

where m_i is the size of the cluster, c_i is the mean of the i th cluster, c is the total mean, and $dist$ is the distance between two objects.

Under the condition that the value of the number of clusters K is determined, K-Means clustering hopes that the smaller the value of the total SSE is, the better it is (the more cohesive the data within the group is), and the larger the value of the total SSB is, the better it is (the more separative the data between the groups is), i.e., the larger the value of the total SSB/total SSE is, the better it is.

In order to eliminate the influence of the number of clusters K and the sample size n on the calculation results, the total SSB/total SSE is corrected to equation (3).

$$\frac{\text{General } SSB}{\text{General } SSE} \bigg/ \frac{n-K}{K-1} \tag{3}$$

where $(n-k)/(k-1)$ is the complexity. Eq. (3) is the Calinski-Harabasz formula, and the larger the ratio, the better.Calinski-Harabasz is simple to compute and fast to run, which will be used in this paper as a reference for determining a reasonable K value. Set the value range of K as 3-10 by enumeration, repeat the run 1000 times on each K value (to avoid local optimal solution), and calculate the Calinski-Harabasz value of the current K value. And then the K corresponding to the largest value of Calinski-Harabasz is taken as the final K.

(2) Improvement of initial point selection

Initial point selection has a great impact on the convergence speed and stability of the algorithm.K-Means clustering algorithm will randomly select K objects from the data as the initial point if there is no specified initial point.

This study adopts the improved method of selecting the distance as far as possible. Firstly, an object is randomly selected as the first initial point, secondly, the object with the farthest distance from the object is selected as the second initial point, and then the object with the largest sum of distance from the first two objects is selected as the third initial point, and so on. Following the same principle until K initial points are selected, the selection process is shown in Figure 1.

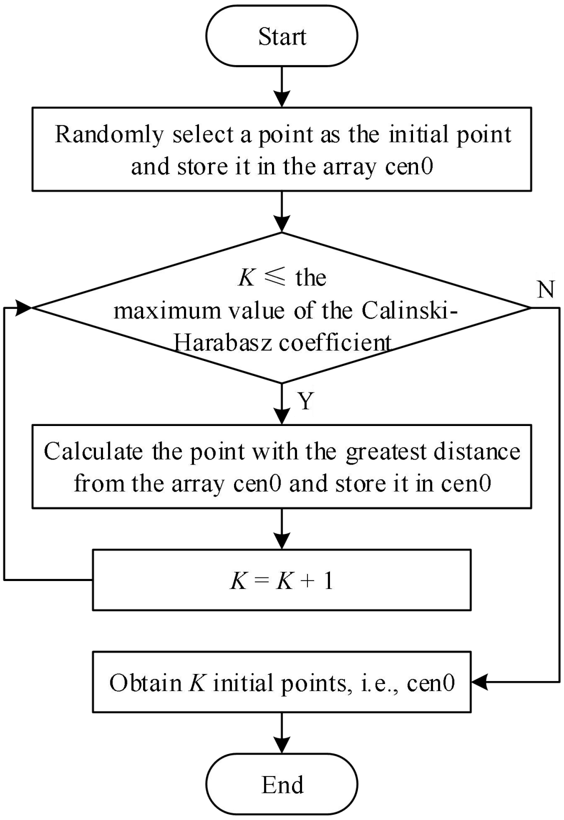


Figure 1. Improved initial point selection.

(3) Evaluation index

The improved K-Means algorithm is now compared with the original K-Mean algorithm for quality and effectiveness. Generally, the clustering evaluation methods are divided into two categories according to whether or not there is a benchmark available: the extrinsic method is used to evaluate if there is an available benchmark, and the intrinsic method is used to evaluate if there is no benchmark available.

In this paper, there is no clustering benchmark, so the intrinsic method is used for evaluation, and the following parametric indicators are used for evaluation:

1) Calinski-Harabasz criterion, see equation (4), the meaning of the formula has been explained earlier and will not be repeated.

2) The mathematical formula for the contour coefficient is as follows:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4}$$

where $a(i)$ denotes the average value of the distance between object i and other objects of the class to which i belongs. The smaller the value, the more compact the inverse cluster is. $b(i)$ denotes the minimum distance from object i to the class to which it does not belong to i. The larger the value, the more separated from other clusters. The contour coefficient takes values between -1 and 1, the larger the value the better the clustering effect.

3) Dunn's index whose mathematical formula is as follows:

$$DVI = \frac{\min\limits_{0<m\neq n<K}\left\{\min\limits_{\substack{\forall x_i \in \Omega_m \\ \forall x_j \in \Omega_n}}\left\{x_i - x_j\right\}\right\}}{\max\limits_{0<m\leq K}\left\{\max\limits_{\forall x_i, x_j \in \Omega_m}\left\{x_i - x_j\right\}\right\}} \tag{5}$$

Dunn's index first calculates the shortest distance between any two classes, and then divides the result by the maximum distance between objects in any class. A larger Dunn's index means better clustering.

2.3 Analysis of clustering results

2.3.1. Clustering results

After analyzing the results of SPSS19.0 clustering, the female group is divided into five categories of students, the male group is divided into three categories of students, each of which presents different characteristics of its physical test indicators. The specific results are analyzed as follows. The final clustering results of the girls' group are shown in Table 1, and the mean values of the scores of the girls' group after clustering are shown in Figure 2.

From Table 1 and the mean value of Figure 2 can be seen, the girls' group in the clustering 3 of the mean value of the achievement of higher results, such students have better physical quality, clustering 5 of the lowest total score mean value of the students, 47.33 points do not pass the line, its in addition to the seated forward bends, 50m running project on the other items on the lower mean value, indicating that the category of endurance, lower extremity explosive power, waist and abdominal strength are insufficient. In addition, the girls' group in cluster 1, cluster 2, and cluster 4 achieved poorer scores in standing long jump, lung capacity, and 800m running events, respectively, and these three different cluster types of students have their own weak items.

Table 1. Enhanced K-means clustering results for the female group.

Project	Clustering 1	Clustering 2	Clustering 3	Clustering 4	Clustering 5
Sit-reach	86	86	88	80	86
Sit-up	65	70	75	68	41
Standing-jumping	67	66	75	36	28
Time-50m	79	83	85	76	78
Vital-capacity	66	19	73	69	28
Time-800m	37	67	75	69	23
Health Score	66.6667	65.1667	78.5000	66.3333	47.3333

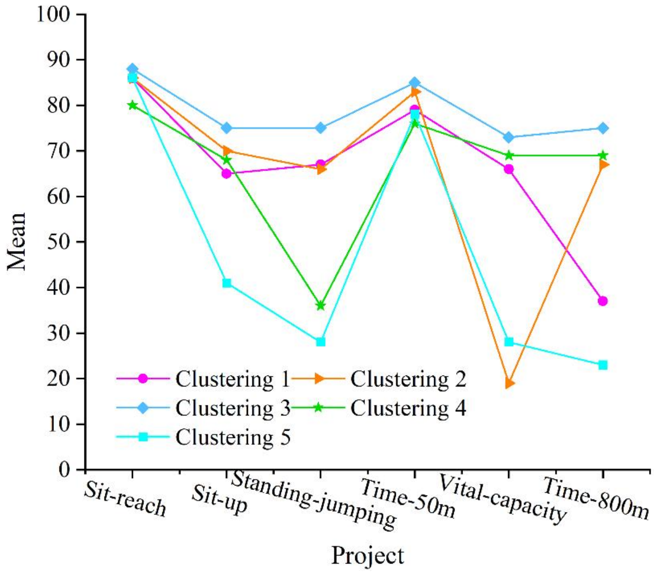


Figure 2. Grouped average scores of female students.

The final clustering results for the boys' group are shown in Table 2, and the mean values of the scores for each item after clustering for the boys' group are shown in Figure 3. From Table 2 and Figure 3, it can be seen that the boys' group had the highest mean values of the scores of the items measured in cluster 1, which showed higher physical fitness in this category, and the boys in cluster 3 showed lower total scores of the test items. Cluster 2 and Cluster 3 both achieved lower scores in the pull-up event and had the highest number of people in both categories, with the lowest mean pull-up score of 4 in Cluster 3. Cluster 3 also achieved lower mean scores in the standing long jump event. It was also found that Cluster 3 students failed in the total mean scores, targeting male students who were mainly weak in the standing long jump and the pull-up to test events.

Table 2. Enhanced K-means clustering results for male group.

	Clustering 1	Clustering 2	Clustering 3
Time-1000m	86	88	83
Time-50m	85	78	72
Standing-jumping	67	62	17
Sit-reach	76	73	64
Pull-up	73	8	5
Vital-capacity	73	64	51
Health Score	76.6667	62.1667	48.6667

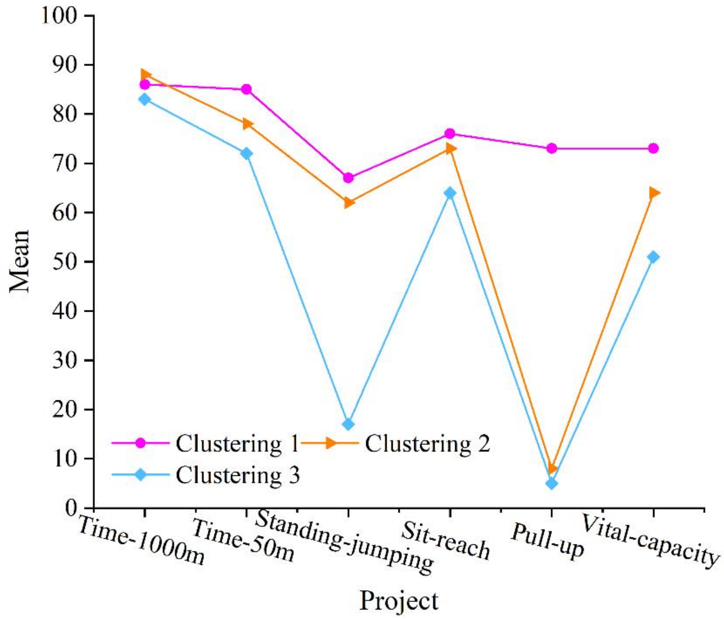


Figure 3. Average scores of male group clusters.

2.3.2. Outlier analysis

The significance of outlier analysis is to find the noise data in the sample data and discover the potential valuable objects. Outlier analysis has been widely used in financial risk control, network intrusion detection, and discovering abnormal climate phenomena. In this study, the purpose of outlier research is to find a small number of students with abnormal physical characteristics. Abnormal physique indicates that the data of students' physique characteristics are quite different from those of the majority of students, and in a practical sense, these students may have exceptionally good physique or exceptionally poor physique, and they are the population that needs to be paid attention to in the future.

There are various methods for analyzing outliers, including density-based, distance-based, distribution-based and clustering-based. In this study, outlier discovery is done by Euclidean distance under the results of improved K-means clustering.

The steps for outlier discovery are as follows:

(1) Calculate the geometric center of mass of the cluster.

(2) Calculate the Euclidean distance of all samples in the cluster from the cluster center.

(3) Visualize the distribution of all points.

(4) Define the small number of samples with the largest distance as outliers.

Based on the medical examination data of this subject, the optimal number of clusters is 3, so only the distribution of outlier points in the case of 3 clusters is analyzed. The clustering result is based on each cluster to calculate the distance of each sample to the cluster center and regarded as the degree of outliers, and the data distribution is visualized, and the result is shown in Figure 4. The figure shows the distribution of the distance of all sample points to the cluster center within 3 clusters, and the point with the largest distance can be subjectively defined as the outlier of the physical examination data. With the visualized results, it can be considered that the Euclidean distances of the outlier points are significantly different from those of the other sample points. The experimental results show that for the 5452 sample data, the combined number of outlier points is 29, which is in line with our basic knowledge of outlier data.

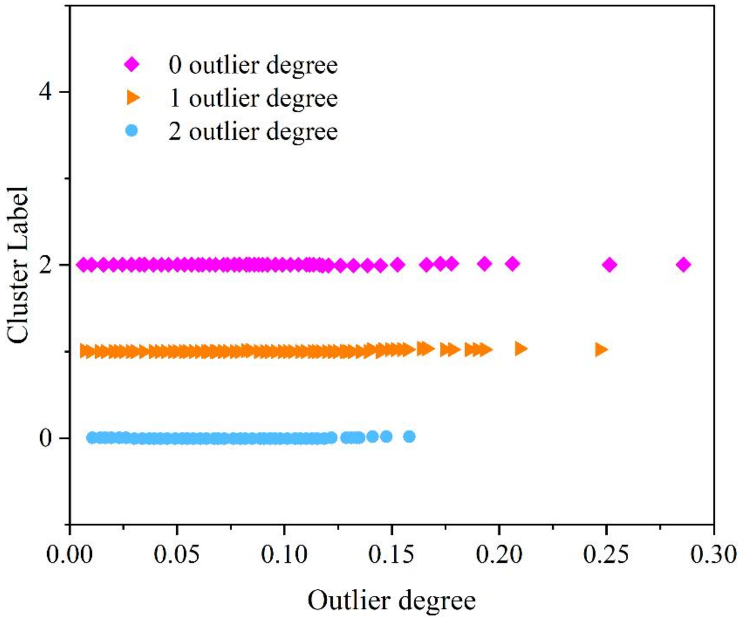


Figure 4. Cluster outlier degree chart.

Table 3 shows the actual data for some of the outliers, showing only the two samples with the greatest degree of outliers in each cluster. The individual students shown in this table can be regarded as objects of special attention. In the actual management of physical education teaching, it is necessary to focus on the sports performance of these students and adjust the content of sports at the right time.

Table 3. The two samples with the greatest cluster outlier degree.

number	sex	vital capacity	stature (cm)	…	Blood pressure-high (mmHg)	Blood pressure-low (mmHg)	Cluster
0123	Female	2512.0	157.2	…	106.0	64.0	0
0147	Female	1543.0	160.0	…	111.0	81.0	0
0163	Male	2512.0	163.0	…	127.0	74.0	1
0182	Male	2743.0	159.0	…	131.0	90.0	1
0199	Female	1600.0	157.0	…	108.0	66.0	2
0207	Female	1400.0	158.0	…	109.0	88.0	3

3. SFD Sport Adaptive Teaching Recommendation Model Based on Body Measurement Big Data

3.1. Algorithm design and data processing

3.1.1. General design of the algorithm

The college student physical test recommendation algorithm is an algorithm designed specifically based on college student physical test big data, and the algorithm flow is shown in Fig. 5. The concept of set-pair analysis is introduced in this study to transform the traditional similarity. In the recommendation process, we first need to calculate the similarity, uncertainty and dissimilarity between the two objects, and then calculate the set-pair recommendation degree $rec(A,B)$ according to the set-pair analysis theory, and finally filter the suitable recommendation objects with the set-pair recommendation degree and make recommendations.

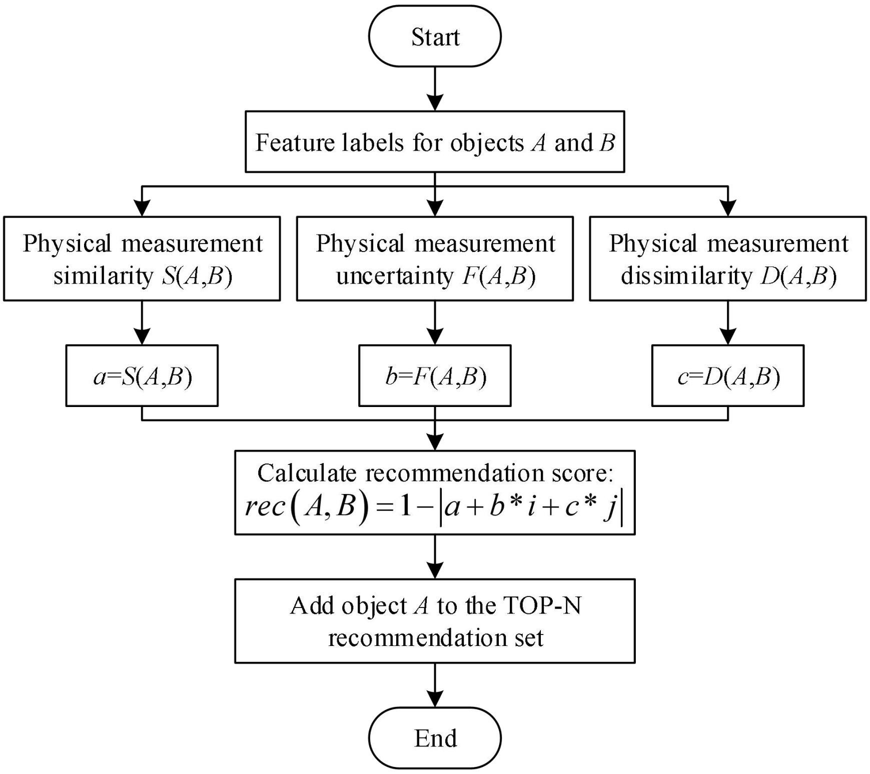


Figure 5. Algorithm flow of physical test for college students.

3.1.2. Data pre-processing

Before the design of the recommendation algorithm, in order to calculate the degree of recommendation of physical testing between 2 objects, it is necessary to score college students' physical testing programs according to the "National Physical Fitness Standard", percentile the data of students' physical testing programs, and classify the students' physical condition: explosive, endurance, flexibility, strength, and different categories of scores are calculated by the students' physical testing program scores according to different weights, and then through the threshold grading to determine the groups of strong, medium and weak students in each category, and finally the literal motor characteristics need to be fitted to give a definition:

Assuming the existence of objects A and B, their motion characteristics are represented as feature vectors, i.e., $A=\langle a_1,a_2,a_3,a_4\rangle$, and $B=\langle b_1,b_2,b_3,b_4\rangle$, and the difference between the features of A and B can be represented by $|a_k-b_k|$ with $k\in\{1,2,3,4\}$.

$|a_k-b_k|=0$, which means that A and B have similarity.

$|a_k-b_k|=1$, means A and B have uncertainty.

$|a_k-b_k|=2$, which indicates that A and B have dissimilarity.

The study will be designed for body measure similarity, dissimilarity, and uncertainty in the context described above.

3.2. Pairwise Recommendation Degree

3.2.1. Similarity of body measurements

Similarity is a numerical measure of the degree of similarity between 2 objects, which is an important reference index in personalized recommendation system, and the traditional similarity is calculated according to the user's rating of the item and the recommendation of the similarity using the relevant formula. The calculation method of similarity of physical measurement in this study is different from the traditional method, which refers to obtaining students' physical measurement data, analyzing and processing the data in order to extract the sports characteristics describing the students, and calculating the similarity by comparing the characteristic values among different students through certain methods. Suppose there are objects A and B in the recommender system, then the following similarity exists between the recommended objects A and B:

$$S(A,B)=\frac{N(a_k=b_k)}{n} \tag{5}$$

where $S(A,B)$ is the similarity of physical measurements between objects A and B, n is the sum of the kinds of characteristic attributes, and a_k and b_k denote the characteristic values of the 2 objects, respectively. By certain rules, the students' performance is divided into different levels, when the 2 students' physical performance levels are equal, they are considered to have certain similarity, i.e., there exists $N(a_k=b_k)$ is the number of features with the same eigenvalues of the 2 objects.

3.2.2. Somatic measurement of dissimilarity

Dissimilarity is a numerical measure that describes the degree of difference between 2 objects, in the recommendation problem based on the physical test scores, it is often unreasonable to recommend only based on the similarity, if the recommended two parties only exist in similarity, it means that the two parties can learn very little from each other, so the two parties need to exist in a certain degree of dissimilarity. Only the existence of dissimilarity between the two sides, there is the possibility of learning from each other. Let there exist 2 objects A and B with dissimilarity in the recommender system, then the following dissimilarity exists in the recommended objects A and B:

$$D(A,B)=\frac{\sum_{k=1}^{n}\left\| a_k-b_k\right|-1\right|-N(a_k=b_k)}{n} \tag{6}$$

Where $D(A,B)$ is the dissimilarity between objects A and B, n is the sum of the types of feature attributes, and a_k and b_k denote the eigenvalues of the 2 objects, respectively. $\sum_{k=1}^{n}\left\| a_k-b_k\right|-1\right|-N(a_k=b_k)$ is the number of dissimilarity features that the 2 objects have, where $\sum_{k=1}^{n}\left\| a_k-b_k\right|-1\right|$ is the sum of the number of similar and dissimilar features that the 2 objects have, $N(a_k=b_k)$ is the total number of similar features of the 2 objects, and n is the kind of feature attributes.

3.2.3. Uncertainty of physical measurements

If there are objects A and B in the recommendation system, when there exists between A and B that one party's sports performance is moderate and the other party's performance is strong or weak, this cannot be used to determine whether the gap between the two parties can really be large enough to teach the other party, and therefore there is uncertainty. In set-pair theory, similarity, dissimilarity and uncertainty sum to 1, that is, $a+b+c=1$, therefore, there exists for the recommended objects A and B there is the following uncertainty:

$$F(A,B)=1-\frac{\sum_{k=1}^{n}\left\| a_k-b_k\right|-1\right|}{n} \tag{7}$$

3.2.4. Determination of the degree of connectedness i

The value of the difference uncertainty coefficient i in the set-pair recommendation degree corresponding to objects A and B in the recommender system is a key point to be determined. When the value of i tends to be more close to 1, the similarity between the 2 objects is higher, therefore, using the cosine similarity, a method of determining the value of i using the computational value method is proposed, which exists in the following definition:

$$i = \frac{1}{1 + d / s} \tag{8}$$

where s is the cosine similarity of objects A, B. The formula is as follows:

$$s = \frac{\sum_{k=1}^{n} (a_k \times b_k)}{\sqrt{a_k}\sqrt{b_k}} \tag{9}$$

d is the cosine dissimilarity of objects A, B. Under certain conditions, the cosine dissimilarity can be transformed from the cosine similarity. The following formula is used to transform the similarity:

$$d = e^{-s} \tag{10}$$

d denotes the dissimilarity between the 2 objects, and s denotes the similarity between the 2 objects; as s takes a larger value, d takes a smaller value, and i tends to be closer to 1; conversely, i is further from 1.

3.2.5. Computation of Top-N Recommendation Set and Friend Recommendation

Top-N recommendation algorithm is to sort the data according to certain rules and select the largest or smallest N data from the sorted list for recommendation. For different social environments, different rules can be formulated to screen the data in the recommendation set, based on the university campus student groups for research, in the friend recommendation at the same time need to consider a variety of factors, according to the user set obtained on the set of the degree of recommendation, calculating the average of the degree of recommendation of the user and the recommended user as a threshold r, will be greater than the threshold degree of recommendation is stored to the user Top-N recommendation set, recommendation process is shown in Figure 6. -N recommendation set, the recommendation process is shown in Figure 6.

Because in the process of friend recommendation, not only need to consider the user and the user's recommendation degree of both, there are also some practical issues need to be considered, such as the user's class, gender, distance from the living area, etc., the recommendation of the user needs to be screened according to the user.

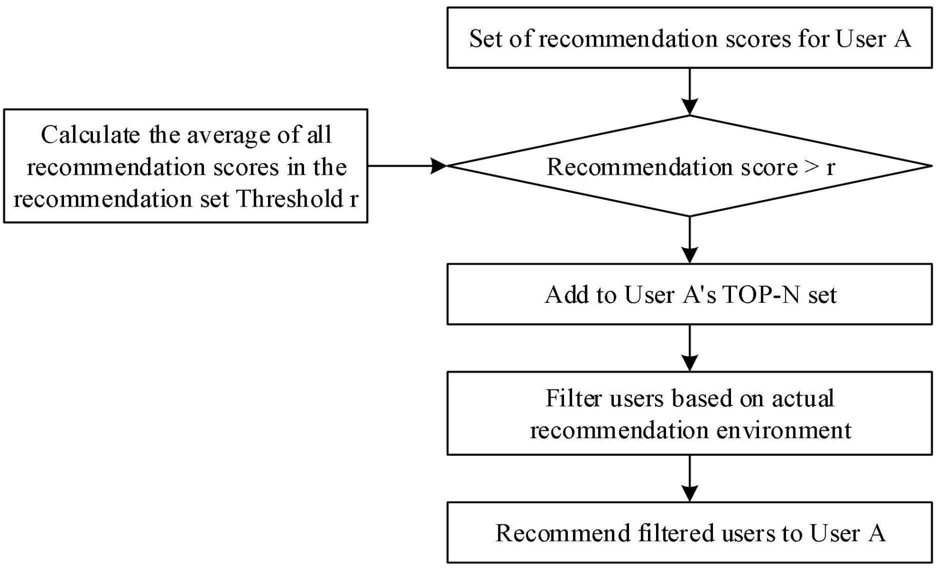


Figure 6. Top-N recommendation set and friend recommendations.

3.3. Results and Analysis of Instructional Campaigns

3.3.1. Experiments and analysis of results

(1) Experimental platform construction

Experimental platform: install ubuntu server 17.10 on the virtual machine, as well as Java 8, Hadoop 2.7, Spark2.3.2, to build the Spark computing cluster platform.

Experimental data: 4 sets of experiments were conducted using the UCI public MovieLens dataset, 100K, 1M, 10M and 100MB corresponding to 100,000, 1,000,000, 10,000,000 and 100,000,000,000 pieces of data respectively. The ratings are categorized into 1 to 5 levels, with larger values indicating that users rate the item more highly.

(2) Rating prediction accuracy test

Root Mean Square Error (RMSE), Mean Absolute Error (MAE) are the most commonly used indicators to measure the accuracy of recommendation system recommendation, the smaller the value of both means the better the recommendation effect, the recommendation results are shown in Table 4. From the data in the table, it can be seen that compared with the traditional collaborative filtering algorithm, the MAE and RMSE values of the friend recommendation model based on set-pair theory are smaller, which indicates that the proposed method can push up the recommendation accuracy.

Table 4. Traditional versus MAE and RMSE values in this study.

Different algorithms	MAE	RMSE
Traditional collaborative filtering algorithm	0.925	0.934
Recommended in this article	0.889	0.893

(3) Comparison of algorithm execution time test

The traditional collaborative filtering algorithm, Spark-based collaborative filtering algorithm and set-pair theory-based buddy recommendation algorithm are respectively compared for different data sets in the experiment, and the results are shown in Figure 7. From the data, it can be seen that with the increase in the amount of processed data, the time efficiency advantage of friend recommendation based on set-pair theory is becoming more and more obvious, providing the possibility of real-time accurate recommendation.

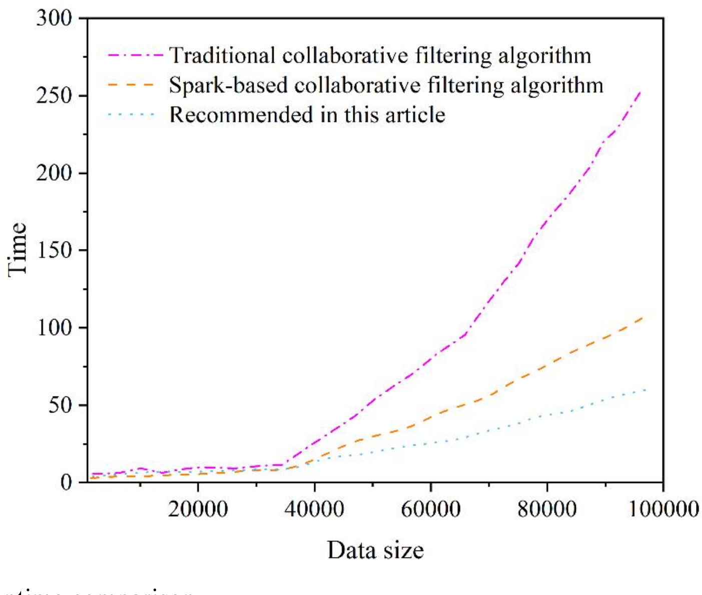


Figure 7. Model runtime comparison.

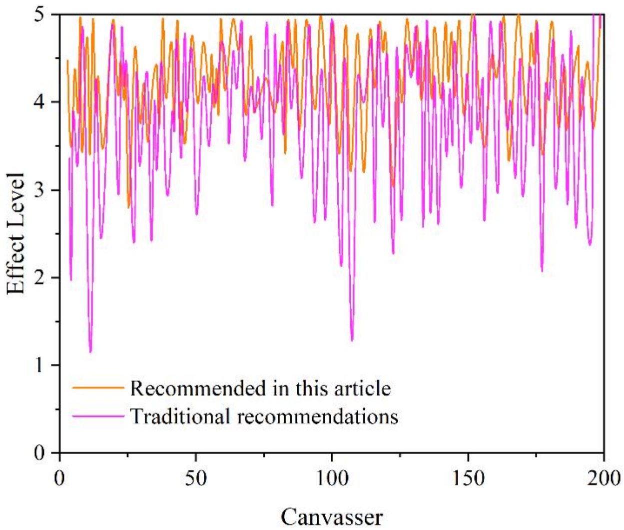
3.3.2. Exercise prescription parameter recommendation experiments

Since the main application scenarios of current recommendation algorithms are mostly in recommending videos, news, commodities, etc., and the recommendations are specific entities, the direct recommendation of entities is generally used, without changing the information of entities. At present, the most commonly used recommendation method is cooperation-based recommendation, i.e., looking for user groups with similar interests and then recommending the information of entities with high interest in the group. According to this recommendation method, exercise prescription direct recommendation means finding similar groups of current exercisers and then recommending the effective prescription information within the group to current exercisers. In this paper, 200 athletes in the test set are randomly selected, and the friend recommendation method based on set-pair theory and the direct recommendation method based on cooperation are utilized to recommend the parameters of exercise prescription for these 200 athletes, and the prescription effect level under the two recommendation methods is obtained by using the prescription effect simulation program. Its Table 5 shows the effect ratings produced by some exercisers under the two recommendation algorithms, and the experimental comparison between this paper's recommendation and traditional recommendation is shown in Figure 8.

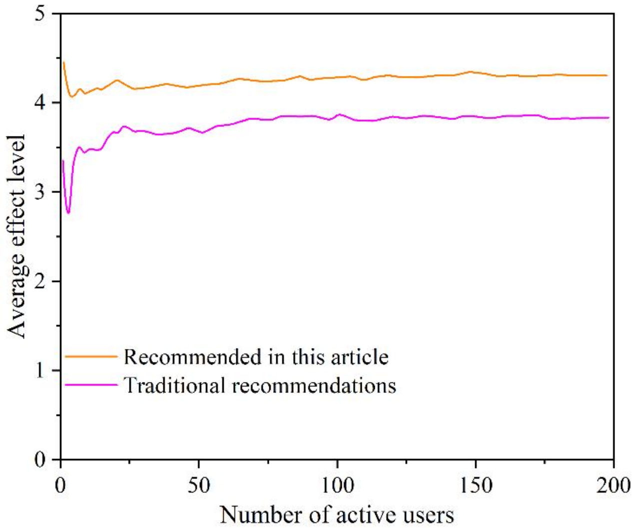
From the experimental results, it can be seen that in terms of exercise prescription recommendation, the overall effect of the buddy recommendation approach based on set-pair theory is better than that of the direct recommendation approach based on cooperation. More importantly, the effect stability of buddy recommendation based on set-pair theory is better, which can effectively ensure that the effect of the recommended exercise prescription will not be too bad, while the direct recommendation based on cooperation is susceptible to the influence of noisy case data, and the stability of the effect is poorer. The stability of the recommended effect in the sports prescription recommendation requires a higher level of stability because, in other scenarios, the user can determine the advantages and disadvantages of the recommendation and the recommended effect does not have much impact on the user, but in the sports prescription recommendation, the user is generally unable to determine the advantages and disadvantages of the sports prescription, and the test period of the recommended effect of the prescription is longer, resulting in a greater impact of the recommended effect on the user. Therefore, it is very important to ensure the stability of exercise prescription recommendation. The experimental results show that the buddy recommendation method based on set-pair theory used in this paper has better overall effect and stability compared to the direct recommendation method.

Table 5. Suggested and traditional recommendations.

Canvasser	1	2	3	4	5	6	7	8	9	10	11	…	200
Recommended in this article	4.6	3.6	4.52	4.33	3.9	5.1	3.7	4.9	3.5	5.6	4.3	…	5.1
Traditional recommendations	3.5	2.1	3.79	3.94	3.1	4.8	3.9	3.6	1.4	4.7	2.6	…	5.3



(a) Effect level comparison chart for 200 athletes



(b) Comparison of the average effect levels of the first i athletes

Figure 8. This article recommends a comparison with traditional recommendation experiments.

4. Experiments with adaptive instructional interventions based on student fitness data

4.1. Selection of experimental subjects

In the process of volunteer recruitment and screening of experimental subjects, volunteers were volunteered to participate in this exercise intervention experiment. People under the age of 20 who were willing to exercise were recruited through on-site enrollment and telephone reservation as experimental subjects. In order to reduce the risk of the experiment and ensure the safety of the subjects, the following conditions were excluded from the screening of the subjects: patients with diseases; patients with lung/nephropathy and other related conditions; patients with severe diabetes or related complications; people with unstable blood pressure or glucose control; and people who could not be fed back and followed up for other reasons.

Considering the operability of the experiment, the research team selected 66 eligible persons as subjects in a middle school. According to the exercise habit or not, we divided the experimental personnel into the exercise habit formation group and the scientific fitness guidance group. There are 32 people in the exercise habit formation group and 34 people in the scientific fitness guidance group.

4.2. Experimental Procedures

The experimental subjects were subjected to a physical health checkup and a comprehensive health questionnaire. The equipment used for the health checkup was the professional non-invasive physical examination equipment developed by our research group in the early stage. The physical examination equipment includes: cardiovascular function detector, arteriosclerosis measuring instrument, ultrasonic bone density detector, and body composition analyzer. The health questionnaires and medical checkups were completed under the guidance and assistance of professional administrators.

During the 8-week exercise intervention experiment, key indicators such as blood pressure, body composition, bone density, and arteriosclerosis degree of the experimental subjects were tracked and tested. Measurements were taken using the same equipment as during the initial physical examination. During the measurement, in order to avoid external interference as much as possible, the temperature of the testing room was kept suitable and similar to that of the initial measurement, and the noise in the testing room was lowered to keep the testers' emotions stable. Measurement of the stability and repeatability of the measurement, each device measured twice, the interval between the two times is 1 minute, if the difference between the two measurements is large, then re-measurement, until the measured data are stable. After the completion of the experiment, the movement data and key index data of the experimental subjects were organized and analyzed.

4.3. Analysis of experimental results

After completing the 8-week exercise intervention trial, the exercise data and key index data of the experimental subjects were organized and analyzed. Due to various reasons, there were four experimental subjects in the exercise habit formation group who did not follow the exercise prescription and the total amount of exercise was not recorded, so all the data of these four subjects were excluded from the data processing and analysis. Therefore, the valid data of the exercise habit cultivation group was 28 cases, and the total experimental subjects were 66 cases. This experiment was calculated using SPASS 20.0 statistical analysis software. The average daily step counts at baseline and during the intervention period for all experimental subjects as well as for the subgroup experimental subjects are shown in Table 6. The baseline weekly exercise totals (METs-min/week) for all subjects as well as grouped subjects and the weekly exercise totals for each week during the intervention period are shown in Table 7.

Calculations using the paired-samples t-test based on the data analysis in Tables 6 and 7 showed that the baseline number of steps of the experiment and the number of steps of each week during the experiment were significantly different (p<0.001), and the baseline exercise volume of the experiment and the exercise volume of each week during the experiment were also significantly different (P<0.001).

Table 6. Exercise data for baseline and intervention weeks.

Intervention Period	0	1	2	3	4	5	6	7	8
All subjects (N=66)	6914±3627	9418±3811	9724±3853	9624±3862	9784±4114	9741±4103	9873±3742	9986±3628	10279±3542
Exercise Habit Cultivation Group (N=32)	4831±1706	7427±3122	7541±2716	7432±2264	7052±2315	7318±2849	7852±2486	8417±2315	8573±2019
Scientific Fitness Guidance Group (N=34)	8814±3743	12057±3014	12404±3729	12392±3674	12843±3781	12615±3462	12663±3854	12432±3643	12544±3361

Note 1: Sunday mean steps are in the form of mean ± standard deviation.
2: A paired samples t-test was used to demonstrate the difference in the average number of steps per day per week, with a significant difference between the data from each week during the intervention period and the baseline data (p<0.001), while there was no significant difference between the data from each week during the intervention period (p>0.05).

Table 7. Total exercise during baseline and intervention weeks(METs-min).

Intervention Period	0	1	2	3	4	5	6	7	8
All subjects (N=66)	407±472	1015±729	1048±719	1073±3733	1077±756	1095±754	1164±867	1135±809	1203±816
Exercise Habit Cultivation Group (N=32)	——	572±419	608±426	643±389	577±388	608±374	663±375	672±384	718±355
Scientific Fitness Guidanc	734±397	1384±634	1401±615	1422±662	1437±639	1492±651	1574±638	1509±669	1595±674

e Group (N=34)									

Note 1: Weekly exercise totals are in the form of mean ± standard deviation.
2: Baseline data for the exercise habit parenting group were not accounted for in the calculations because there was no regular exercise.
3: The difference in weekly exercise totals was demonstrated with a paired samples t-test, and there was a significant difference between the data from each week during the intervention period and the baseline data ($p<0.001$), while there was no significant difference between the data from each week during the intervention period ($p>0.05$).

The health indicators of all subjects before and after the experiment are shown in Table 8. From the data analysis in the table, it was calculated that the difference between the BMI, SEVR, DBP, SBP, CAP, and CP indexes of all subjects before and after the intervention experiment was very significant ($p<0.001$), with a large improvement; moreover, it was found by comparative analysis that there was a significant difference between the muscle ratio, fat ratio, abdominal fat volume, and PP indexes before and after the experiment ($p<0.05$), with a certain improvement.

Table 8. Health indicators of all subjects before and after the experiment.

Metric	Before the experiment		Post-test		P
	Average value	Standard deviation	Average value	Standard deviation	
BMI (kg/m²)	25.17	3.24	24.81	3.23	<0.001
Muscle ratio (%)	72.54	7.18	74.09	6.65	0.031
Fat percentage (%)	25.59	7.29	24.71	6.83	0.032
Abdominal fat mass (kg)	9.34	3.62	9.03	3.55	0.009
HR	72.64	8.19	70.84	7.19	0.089
SEVR	1.07	0.23	1.26	0.24	<0.001
DBP (mmHg)	77.43	9.59	72.47	8.93	<0.001
SBP (mmHg)	123.41	13.14	115.28	13.29	<0.001
CAP (mmHg)	109.28	17.08	103.19	16.47	<0.001
PP (mmHg)	47.09	7.91	44.13	7.08	0.009
ABI (right)	1.18	0.09	1.22	0.75	0.273
ABI (left)	1.19	0.09	1.21	0.09	0.882
ABI (mean)	1.18	0.07	1.22	0.08	0.359
PWV (right)(m/s)	13.27	1.59	13.15	1.64	0.542
PWV (left) (m/s)	13.40	1.88	13.09	1.67	0.772
PWV (mean) (m/s)	13.52	1.64	13.10	1.58	0.801
STI	91.43	21.37	88.94	23.47	0.255
T value	-0.87	1.15	-0.99	1.15	0.215
Z value	-0.35	1.26	-0.54	1.33	0.161

The health indicators of all experimental subjects, the exercise habit breeding group and the scientific fitness guidance group after the experiment are shown in Table 9. From the data analysis in the table, it was calculated that the difference between the SBP and CAP indicators of the exercise habit breeding group before and after the intervention experiment was very significant ($P<0.001$), with a large improvement; moreover, it was found that there was a significant difference between the BMI, DBP, and SEVR indicators of the pre-experiment and post-experiment through comparative analysis ($P<0.05$), with a certain degree of improvement.

Table 9. Health indicators of exercise habit formation group before and after the experiment.

Metric	Before the experiment		Post-test		P
	Average value	Standard deviation	Average value	Standard deviation	
BMI (kg/m²)	25.07	2.74	24.50	0.81	0.017
Muscle ratio (%)	71.14	7.15	71.29	5.28	0.315
Fat percentage (%)	26.09	7.34	26.53	5.33	0.309
Abdominal fat mass	9.25	3.12	9.14	3.12	0.338

(kg)					
HR	71.56	8.24	69.76	7.01	0.475
SEVR	1.06	0.26	1.19	0.21	0.022
DBP (mmHg)	77.54	9.67	72.17	9.22	0.003
SBP (mmHg)	122.73	13.01	113.62	13.49	<0.001
CAP (mmHg)	109.75	17.78	103.29	16.51	<0.001
PP (mmHg)	47.11	8.74	44.29	7.61	0.087
ABI (left)	1.20	0.12	1.25	0.09	0.099
ABI (right)	1.23	0.12	1.25	0.08	0.788
ABI (mean)	1.19	0.11	1.25	0.08	0.194
PWV (left) (m/s)	14.17	2.09	13.25	1.85	0.534
PWV (right) (m/s)	14.03	1.85	13.13	1.79	0.116
PWV (mean) (m/s)	14.53	1.89	13.22	1.77	0.095
STI	95.16	22.84	89.28	24.37	0.133
T value	-0.67	1.25	-0.93	1.25	0.126
Z value	-0.15	1.36	-0.52	1.28	0.075

The health indicators of the scientific fitness guidance group before and after the experiment are shown in Table 10. From the data analysis in the table, it was calculated that the difference between the BMI, abdominal fat mass, SEVR, DBP and SBP indicators of the scientific fitness guidance group before and after the intervention experiment was very significant ($P<0.001$), with a large improvement; in addition, the muscle ratio, fat ratio, CAP and PP indicators before and after the experiment were found to be significantly different ($P<0.05$) through comparative analysis, with a certain degree of improvement.

Table 10. Health indicators of scientific fitness guidance group before and after the experiment.

Metric	Before the experiment		Post-test		P
	Average value	Standard deviation	Average value	Standard deviation	
BMI (kg/m²)	25.54	3.84	25.12	3.64	<0.001
Muscle ratio (%)	72.82	7.16	74.59	7.35	0.042
Fat percentage (%)	24.39	7.39	23.51	7.53	0.039
Abdominal fat mass (kg)	8.87	3.92	9.09	4.06	<0.001
HR	71.79	8.05	69.56	7.49	0.077
SEVR	1.14	0.19	1.25	0.26	<0.001
DBP (mmHg)	76.84	9.54	71.22	8.65	<0.001
SBP (mmHg)	122.03	13.54	112.16	13.21	<0.001
CAP (mmHg)	108.15	17.28	103.09	16.13	0.002
PP (mmHg)	46.21	7.74	43.51	6.51	0.054
ABI (left)	1.19	0.09	1.19	0.08	0.729
ABI (right)	1.19	0.09	1.19	0.07	0.703
ABI (mean)	1.19	0.09	1.19	0.07	0.985
PWV (left) (m/s)	13.81	1.79	13.76	1.49	0.846
PWV (right) (m/s)	13.73	1.55	13.82	1.55	0.647
PWV (mean) (m/s)	13.86	1.51	13.82	1.53	0.681
STI	89.26	20.24	88.91	23.74	0.884
T value	-1.07	1.05	-1.03	1.26	0.909
Z value	-0.51	1.06	-0.54	1.26	0.891

In summary, after the adaptive teaching intervention experiment can effectively improve the health indicators of students and improve the physical treatment has a certain role.

5. Conclusion

(1) The study takes the physical fitness test data of undergraduates from grade 2023 to 2024 as an example, and proposes an improved K-Means algorithm, and from the results of the clustering analysis, it can be seen that the clustering method based on the improved K-means algorithm can be used to Students are quickly classified according to physical fitness categories, and male and female groups show greater differences in physical fitness influencing factors, and then based on the results of the clustering of

outlier analysis, for the school to guide the physical exercise as a reference.

(2) Proposed a friend recommendation model based on set-pair theory, and the algorithm was evaluated and compared, and the results showed that this paper's recommendation algorithm time efficiency advantage is obvious can significantly improve the recommendation effect. Finally, an exercise intervention experiment was conducted, and the experimental data were statistically analyzed, which concluded that this paper's recommendation system has a certain effect on improving students' health indicators and physical therapy.

References

1. Kyrychenko, V., Deak, G. F., Pop, N. H., & Gomboş, L. (2023). Features of physical development and physical fitness of students from the Faculty of Physical Education and Sport, Babeş-Bolyai University. Journal of Physical Education and Sport, 23(2), 510-516.
2. Keating, X., Liu, X., Stephenson, R., Guan, J., & Hodges, M. (2020). Student health-related fitness testing in school-based physical education: Strategies for student self-testing using technology. European Physical Education Review, 26(2), 552-570.
3. Gula, L. P. (2024). Health and skill-related physical fitness tests: An assessment in determining the students' physical fitness level. Retos, 58, 1014-1021.
4. Simonton, K. L., Mercier, K., & Garn, A. C. (2019). Do fitness test performances predict students' attitudes and emotions toward physical education?. Physical Education and Sport Pedagogy, 24(6), 549-564.
5. Griban, G. P., Lyakhova, N., Tymoshenko, O., Domina, Z., Dovgan, N., Kruk, M., ... & Prontenko, K. (2020). Current state of students' health and its improvement in the process of physical education. Wiadomości Lekarskie, (73 (7)), 1438-1447.
6. Mamuti, F. (2019). Transition of Physical Education from Traditional to Modern System, based on The European Credit Transfer System (ECTS). SPORT AND HEALTH International Journal of Sport Sciences and Health, 6(11-12), 54-62.
7. Philpot, R., & Smith, W. (2018). Making a different difference: Physical education teacher education students' reading of critical PETE program. Curriculum Studies in Health and Physical Education, 9(1), 7-21.
8. Colquitt, G., Pritchard, T., Johnson, C., & McCollum, S. (2017). Differentiating instruction in physical education: Personalization of learning. Journal of Physical Education, Recreation & Dance, 88(7), 44-50.
9. Kirch, A., Schnitzius, M., Spengler, S., Blaschke, S., & Mess, F. (2021). Knowing students' characteristics: Opportunities to adapt physical education teaching. Frontiers in Psychology, 12, 619944.
10. Quennerstedt, M., Barker, D., Johansson, A., & Korp, P. (2025). Teaching with the test: Using fitness tests to teach paradoxically in physical education. European Physical Education Review, 31(3), 462-481.
11. Ogata, H., Majumdar, R., Flanagan, B., & Kuromiya, H. (2024). Learning analytics and evidence-based K12 education in Japan: usage of data-driven services for mobile learning across two years. International Journal of Mobile Learning and Organisation, 18(1), 15-48.
12. Sajja, R., Sermet, Y., Cwiertny, D., & Demir, I. (2025). Integrating AI and learning analytics for data-driven pedagogical decisions and personalized interventions in education. Technology, knowledge and learning, 1-31.
13. Zhao, J., & Liu, X. (2024, October). Research on Precision Teaching in Higher Education Based on Big Data. In 2024 10th International Conference on Big Data and Information Analytics (BigDIA) (pp. 1-4). IEEE.
14. Rafiq, S., Zaki, K. A., & Nawaz, A. (2025). Personalized Learning in the Digital Age: Harnessing Technology for Student Success. Contemporary Journal of Social Science Review, 3(1), 1515-1528.
15. Azhyppo, O., Pavlenko, V., Mulyk, V., Mulyk, K., Karpets, L., Grynova, T., & Sannikova, M. (2018). Direction of teaching the subject of physical education by taking into account opportunities of institution of higher education and interests of student youth. Journal of Physical Education and Sport, 18(1), 222-229.
16. Ali, N., Chen, J., Fu, X., Hussain, W., Ali, M., Iqbal, S. M., ... & Thanh, H. V. (2023). Classification of reservoir quality using unsupervised machine learning and cluster analysis: Example from Kadanwari gas field, SE Pakistan. Geosystems and Geoenvironment, 2(1), 100123.
17. Maechler, M. (2019). Finding groups in data: Cluster analysis extended Rousseeuw et al. R package version, 2(0), 242-248.
18. Liu, R. (2022). Data analysis of educational evaluation using K-means clustering method. Computational Intelligence and Neuroscience, 2022(1), 3762431.
19. Burns, R., Fu, Y., Brusseau, T., & Constantino, N. (2017). A Cluster Analysis and Validation of Health-Related Fitness Tests in College Students. J. Phys. Act. Res, 2, 73-79.
20. Gao, X., Rosas, M. F., Zhang, Y., Gatpandan, M. P., Gatpandan, P. H., & Dasig, D. (2021, July). The Research and Analysis of Student Health Status on the Basis of Physic Fitness Test'Data. In 2021 International Conference on Public Health and Data Science (ICPHDS) (pp. 110-115). IEEE.
21. Cepkova, A., Cepka, R., Šooš, L., Uvacek, M., Žiška, J., Honz, O., & Zemková, E. (2025). Cluster analysis as an effective tool for identifying physical fitness in students: the basis for an innovative approach to optimizing physical education in the university environment. Frontiers in Physiology, 16, 1634125.
22. Gao, S., Lu, M., & Miao, N. (2018, April). Physical fitness clustering analysis based on self-organizing feature maps network. In 2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC) (pp. 261-264). IEEE.

23. Hart, P. D. (2017). Profiling physical fitness attributes in college students: A cluster analysis. International Journal of Physiology, Nutrition, and Physical Education, 2(2), 741-744.
24. Tang, P., Wang, Y., & Shen, N. (2020). Prediction of college students' physical fitness based on K-means clustering and SVR. Computer Systems Science and Engineering, 35(4), 237-246.
25. Zhang, L., & Jiao, Y. (2023, October). Analysis and Research on College Students' Physical Fitness Test Data Based on K-means Clustering Algorithm. In Proceedings of the 2023 4th International Conference on Machine Learning and Computer Application (pp. 454-460).
26. Ji, R., Yang, J., Wu, Y., Li, Y., Li, R., Chen, J., & Yang, J. (2024). Construction and analysis of students' physical health portrait based on principal component analysis improved Canopy-K-means algorithm. Journal of Supercomputing, 80(11).
27. Feak, G. F., Kyrychenko, V., Pop, N. H., Khurtenko, O., Zhovtenko, Y., Pidhaina, V., & Boros-Balint, I. (2025). Clustering undergraduate students enrolled in physical education majors in Romania and Ukraine based on their physical fitness levels. Journal of Physical Education and Sport, 25(1), 167-174.
28. Guo, R., Dong, R., Lu, N., Yu, L., Chen, C., Che, Y., ... & Yang, J. (2025). Physical Health Portrait and Intervention Strategy of College Students Based on Multivariate Cluster Analysis and Machine Learning. Applied Sciences, 15(9), 4940.
29. Yang, Y. (2021, September). Evaluation model and application of college students' physical fitness based on clustering extraction algorithm. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 547-552).
30. Yan, J. (2025, May). Research on College Physical Fitness and Health Testing Combining Clustering Algorithm and Association Rule Mining. In 2025 2nd International Conference on Intelligent Computing and Robotics (ICICR) (pp. 960-963). IEEE.
31. Tong, J., & Wang, F. (2022, June). Fuzzy Clustering Prediction and Improvement Path Analysis of University Students' Physique. In 2021 International conference on Smart Technologies and Systems for Internet of Things (STS-IOT 2021) (pp. 96-100). Atlantis Press.
32. Zeng, W., Liu, Y., Li, R., & Zeng, Z. (2021, May). Research on the application of big data in the reform of college education mode—taking sports as an example. In 2021 2nd International Conference on Computers, Information Processing and Advanced Education (pp. 1344-1351).
33. Naomi O. Riches, Ramkiran Gouripeddi, Robert M. Silver & Julio C. Facelli. (2025). Air pollution mixtures and stillbirth: k-means cluster analysis. Hygiene and Environmental Health Advances,16,100153-100153.https://doi.org/10.1016/J.HEHA.2025.100153.