Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7393986, 7 pages
https://doi.org/10.1155/2022/7393986

Research Article

Analysis of Physical Test Indexes of College Students Based on Data Mining Model

Junwu Suo,[1] **Cuixiang Guo**,[2] **and Guifang Wang**[1]

[1]*Shandong Jianzhu University, Jinan 250101, China*
[2]*Shandong Polytechnic, Jinan 250104, China*

Correspondence should be addressed to Cuixiang Guo; guocuixiang@sdp.edu.cn

Received 21 September 2021; Accepted 26 March 2022; Published 14 April 2022

Academic Editor: Gengxin Sun

This paper takes the physical fitness test data and the physical health self-assessment data as the research objects. The decision tree algorithm is used to construct a decision tree model for students who fail to meet the physical test. Thus, the classification of students with different physical qualities is realized. The association rule Apriori algorithm is used to mine the association of physical fitness test indexes so as to judge the hidden law between students' physical fitness and behavior habits and get the correlation information of various physical health indexes. The back propagation (BP) neural network algorithm is used to establish the physical fitness test prediction model. By using these data mining models, this paper explores the hidden association information in college students' physical test data, which can provide more scientific and effective guidance for students' physical tests.

1. Introduction

There are many reasons for the decline of young people's physical health, one of which is that young people do not cultivate good physical exercise habits and do not use scientific methods for physical exercise. At present, exercise has become an important measure to improve the level of physical health in many countries. Maintaining an appropriate amount of exercise every day is beneficial to the health of people of all ages. Exercise can effectively enhance physical fitness and improve psychological quality.

For a long time, colleges and universities paid too much attention to the factors of college students' intellectual development and ignored their physical condition. With the improvement of people's daily living standards, the material life of college students has become richer, resulting in a lack of exercise and the physical quality of some students has decreased. The unsatisfactory physical quality will not only affect the studies and lives of college students but also affect their working state towards the society in future.

Physical fitness tests could make college students deeply realize the importance of physical exercise and urge them to strengthen exercise to improve their physical health. The physical fitness index of college students is the urgent need and development trend of national physical quality, and it is a powerful guarantee for daily study and life. With the continuous development of data mining technology and the wide application of database management systems, colleges and universities have accumulated a large number of college students' physical index data, and the amount of data will continue to grow. Faced with more and more rapidly expanding data, colleges and universities are often unable to understand information, making it difficult for traditional information technology and tools to obtain valuable knowledge.

The traditional statistical analysis of physical fitness test data often adopts a manual method. For teachers, it not only increases the daily workload but also is difficult to effectively process and analyze the data. For students, it is impossible to get real-time and effective feedback from daily teaching activities and physical fitness tests. From the perspective of analysis methods, the traditional analysis methods still stay in the simple calculation of variance, mean, and reliability, resulting in the conclusion of the analysis staying on the

surface, and being unable to give full play to the full value of a large number of data.

With the in-depth study of data mining technology, the application of data mining has gradually expanded to various fields. Some scholars had applied data mining technology to the education field. Baker et al. [1] discussed the importance of applying data mining technology to education and teaching systematically. Kangaiammal et al. [2] used rough set theory to mine and analyze the correlation between the improvement of students' performance and the leading factors of active learning. Fang et al. [3] used the classification method of a decision tree, applied data mining technology to student achievement information, and constructed the professional ability decision tree model. Yadav et al. [4] adopted the decision tree ID3 algorithm and the association rule Apriori algorithm for data mining analysis based on student achievement data. Through the analysis of the ID3 algorithm, what factors are related to students' excellent performance can be obtained. Through the analysis of Apriori algorithm, the influence of the excellence of a course on other courses is mined. Chunduri et al. [5] used the FP-growth algorithm to study students' physical health test data, the results show that nearly half of the students' weight does not meet the standard, and through the operation results of the algorithm, it is observed that the students lack the training of lower limb strength in physical training, and the vital capacity grade and endurance grade are obviously weak. Bishop et al. [6] used the association rule Apriori algorithm to screen out five strong association rules about males and females in college students' physical test data. The results show that more females fail in the standing long jump and more males fail in the pull-up. Ramanathan et al. [7] recorded the data of students' arm strength by using the ID3 algorithm and scientifically analyzed the scientific methods to improve their arm strength. Maldonado-Erazo et al. [8] have customized scientific sports programs and efficient sports plans for athletes by using data mining and comprehensive analysis. In order to predict sports performance more accurately, Xu et al. [9] proposed a hybrid prediction system based on a genetic algorithm and an artificial neural network. The hybrid prediction system is also used to study the impact of physical education on the motor and language ability, social skills, and work ability of mentally retarded children, and to predict the social adaptability of different research objects after a sports training cycle. Jin et al. [10] used a cluster analysis algorithm and a neural network model to evaluate the sports' ability and performance of aerobics athletes, selected appropriate evaluation indicators, and established an evaluation model reflecting competitive ability and performance, which reflected the relationship more scientifically between sports ability and performance.

Based on the physical fitness test data of college students, this paper observes the actual physical fitness of college students through data mining algorithms. These algorithms can obtain more valuable hidden information to help students more clearly understand their self-physical ability and help teachers carry out corresponding teaching activities in time.

2. Analysis of Students' Physical Fitness Data Based on Decision Tree Model

In data mining algorithms, the decision tree model [11] is a particularly common algorithm. It can not only deal with classification problems but also solve regression problems. By learning from the data, the decision tree algorithm classifies and predicts when the values of input variables and output variables are different in different situations. A decision tree is a set of decisions represented by a tree structure. The tree model includes root node, leaf node, and nonleaf node. Leaf nodes are generally represented as a category and connected to the root node are represented as a classification, and the path connected between the root node and leaf node is represented as a classification rule of an attribute.

In the forming process of a decision tree, the main components are feature selection, decision tree generation, and pruning. The key to the decision tree is to select the optimal partition attribute at each split node. With the partition process, make the samples contained in the branch nodes belong to the same category as much as possible. The decision tree algorithm first finds the appropriate features as the root node of the classification according to the actual data samples, and then judges and classifies all other attributes in the data set in turn according to the classification standard of the root node. If the data attribute to be classified is the same as its node attribute, it is divided into the same category. If it is different, it is divided into another category as a new node.

There are a few cases of applying the decision tree to the study of college students' sports health. In this paper, we attempt to mine, analyze and predict college students' sports health data by using the C4.5 algorithm [12]. In order to mine the relationship between students' physical fitness data and students' personal behavior habits data, we first need to calculate which attribute has the greatest correlation between the two data, and classify it as the root node again, so as to form a complete decision tree.

The C4.5 algorithm uses the information gain ratio of attributes as the standard for node selection and considers attributes with a high information gain ratio as "good" attributes. By calculating the information gain ratio of each attribute, the attribute with the highest information gain ratio is selected as the division standard for each division, and this process is repeated many times until a decision tree that can perfectly classify training samples is generated. In the actual operation process, the test attribute is determined by the attribute with the highest information gain ratio. Different test attributes can distinguish multiple different sample subsets. After distinguishing the sample subsets, the stability is evaluated according to the information gain value. If the value of the information gain ratio is low, the instability is large,

if the value of the information gain ratio is high, the instability is small.

In the C4.5 algorithm, set S have s sample data, and there are m category attributes $C_i\,(i = 1, 2, \ldots, m)$, suppose that the number of samples in class C_i is s_j. The total information entropy for this sample is

$$I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2(p_i), \tag{1}$$

where p_i represents the probability that the sample belongs to C_i, it can be defined as follows:

$$p_i = \frac{s_i}{s}, \tag{2}$$

where s represents the total of all samples.

Suppose that the test attribute is A, and A contains k different values $\{a_1, a_2, \ldots, a_k\}$, then set S can be divided into k subsets $\{S_1, S_2, \ldots, S_k\}$, and all samples whose value is equal to a_j belong to subset S_j, which is a subset of the generated new leaf nodes.

Suppose that the number of samples with category C_i in subset S_j is s_{ij}, then the information entropy of the divided samples is

$$E(A) = \sum_{j=1}^{k} \frac{s_{1j} + s_{2j} + \cdots + s_{mj}}{s} I\left(s_{1j}, s_{2j}, \ldots, s_{mj}\right), \tag{3}$$

where $I(s_{1j}, s_{2j}, \ldots, s_{mj})$ and p_{ij} represents probability of samples with category attribute C_i in subset S_j.

Finally, the information gain of sample set S is obtained as follows:

$$\text{Gain}(A) = I(s_1, s_2, \ldots, s_m) - E(A). \tag{4}$$

According to the above formula, it can be seen that the information $\text{Gain}(A)$ increases with the decrease of information entropy $E(A)$, then the uncertainty of classifying the set with A as the test attribute would be reduced k. Different attributes A in the set correspond to k subsets in the set S. The above steps are iterated repeatedly, so as to generate other attributes as the child nodes of the node, and finally, a complete decision tree is constructed.

The split information of attribute A can be defined as

$$\text{Split}I(A) = -\sum_{j=1}^{m} \frac{|S_j|}{|S|} \times \log_2 \frac{|S_j|}{|S|}, \tag{5}$$

where m represents the number of the categories.

2.1. Definition of Information Gain Ratio in Sample Set

$$\text{GainRatio}(A) = \frac{\text{Gain}(A)}{\text{Split}(A)}. \tag{6}$$

This paper selects four attributes related to college students' cardiopulmonary function: diet (DI), work and rest (WR), exercise habits (EH), and attention to physical health (AP), and uses the decision tree C4.5 algorithm to classify the data set. Based on the data information obtained after data preprocessing, the "cardiopulmonary function" attribute in the physical health evaluation table of college students has two different classification values: CF1 and CF2, representing qualified and unqualified, respectively. There are 1137 CF1 samples and 1872 CF2 samples in the data set.

The calculation results of the decision tree C4.5 algorithm combined with physical fitness data are shown in Table 1.

According to the information gain ratio of each attribute in Table 1, the gain ratio of the attribute "exercise habit" in the physical health evaluation of students is the largest. Then, the attribute EH is taken as the root node, branches would be generated according to the value of the attribute, and the two branches are divided into new nodes. The above steps are executed recursively until they cannot be divided, so as to obtain a decision tree on the health status of "cardiopulmonary function" of college students' physical health.

According to the decision tree after pruning, the following classification is obtained:

If "EH" = qualified, then cardiopulmonary function meets the standard.

If "EH" = unqualified and "WR" = qualified, then cardiopulmonary function meets the standard.

If "EH" = unqualified, "WR" = unqualified, "DI" = qualified, and "AP" = qualified, then cardiopulmonary function meets the standard.

If "EH" = unqualified, "WR" = unqualified, "DI" = qualified, and "AP" = unqualified, then cardiopulmonary function does not meet the standard.

If "EH" = unqualified, "WR" = unqualified, "DI" = unqualified, and "AP" = qualified, then Cardiopulmonary function meets the standard.

If "EH" = unqualified, "WR" = unqualified, "DI" = unqualified, and "AP" = unqualified, then cardiopulmonary function does not meet the standard.

Based on the above classification results, it is concluded that among the four indexes of college students' cardiopulmonary function, it has the greatest correlation with "exercise habit". Therefore, students with good exercise habits often exercise their cardiopulmonary function. Compared with students with less exercise habits, their cardiopulmonary function will be strengthened.

Students' work, rest, and diet are also related to the strength of their cardiopulmonary function. A good work-and-rest relationship will greatly enhance human immunity and health, and the system will continue to decline for students who stay up late. The diet may not be healthy and even lead to obesity, which will greatly reduce the cardiopulmonary capacity.

The importance of physical health is related to cardiopulmonary function. Actively participating in classroom and extracurricular sports activities and independently and developing scientific sports habits can enhance physical fitness.

3. Analysis of Students' Physical Fitness Data Based on the Apriori Algorithm

Association rules are used to find frequent patterns between items or objects. It can reveal the internal relationships

Table 1: The calculation results of the decision tree C4.5 algorithm combined with physical fitness data.

Index	DI	WR	EH	AP
E	0.9527	0.9216	0.8699	0.9502
Gain	0.0043	0.0368	0.0816	0.0077
SplitI	0.8687	0.7612	0.9268	0.9986
GainRatio	0.0056	0.0483	0.0882	0.0072

between items or objects. Association rule mining is a rule-based machine learning algorithm, which can find interesting relationships in large dataset. The strength of these relationships usually needs to be measured by three indicators of association rules: support, confidence, and lift [13].

Support generally indicates the popularity of an item set, which is measured by the proportion of the occurrence times of the item set. For example, there are items A and B in set C, and the support of item A is the probability of item A in set C. The support of itemsets A and itemset B is the probability that they occur at the same time.

$$\begin{aligned} \text{Support}(A) &= P(A), \\ \text{Support}(A \Rightarrow B) &= P(A \cap B). \end{aligned} \tag{7}$$

Confidence refers to the probability that itemset B also exists when itemset A exists.

$$\text{Confidence}(A \Rightarrow B) = P(B \mid A). \tag{8}$$

There is a possibility that the confidence may mislead the correlation degree. It is assumed that the confidence value of itemset AB is very high because the support of itemset B is very high. But in fact, it cannot explain that AB is highly related. Therefore, in order to illustrate the basic penetration of the two itemsets, the lift index was introduced.

Lift indicates that itemset A exists and itemset B also exists, but at the same time, the popularity of itemset B should be controlled.

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Support}(A, B)}{\text{Support}(A) \times \text{Support}(B)}. \tag{9}$$

If Lift = 1, it indicates that there is no association between itemsets A and B, if Lift >1, it means that when A exists, B is likely to exist at the same time, if Lift <1, it means that when A exists, B has a high probability that it does not exist.

In the practical application of association rules, the minimum support is used to measure the support threshold of items or objects. At the same time, the minimum confidence threshold used to measure the confidence also represents the minimum reliability standard of the association rules.

Frequent itemset mining is the basis of many important data mining tasks, such as association rules, correlation analysis, and causality. The FP-Growth algorithm [14] uses the idea of divide and conquer to recursively divide the transaction data set into multiple smaller conditional transaction data sets to mine frequent itemsets. The FP-Growth algorithm only needs to traverse the transaction data set twice during execution and does not generate candidate sets, which is one order of magnitude faster than the Apriori algorithm in performance.

In this paper, the Apriori algorithm and the FP-Growth algorithm are used to analyze the correlation of students' physical fitness data and determine the minimum support (min_sup = 50%) of association rules based on the indicators of data. The association rules are determined with the minimum confidence (min_conf = 70%).

The five indexes in the physical test data (cardiopulmonary function, muscle strength, muscle endurance, softness, and obesity) were taken as an example. Students were randomly selected from the data set as research samples, so as to obtain the transaction database. The transaction database D is shown in Table 2.

By scanning all items in transaction database D, candidate 1-itemset C_1 is obtained and frequent 1- itemset L_1 is generated by filtering with the minimum support. The frequent 1-itemset L_1 is connected itself to generate candidate 2-itemset C_2 and in the same way, frequent 2-itemset L_2 is generated by filtering with the minimum support. Then the frequent 2-itemset L_2 is connected itself to generate candidate 3-itemset C_3, and finally, frequent 3-itemset L_3 is generated by pruning.

The calculation results of association rule confidence are shown in Table 3.

According to the confidence shown in Table 3, two relatively stable association rules are obtained: $(I_2, I_4) \Rightarrow I_5$ and $(I_4, I_5) \Rightarrow I_2$. Through the strong association rules, it can be seen that students whose obesity and cardiopulmonary function do not meet the standard are likely to have lower softness, and students whose muscle strength and muscle endurance do not meet the standard are likely to have poor cardiopulmonary function.

4. Students' Physical Fitness Test Prediction Model Based on BP Neural Network

As a part of a neural network, the BP neural network is a supervised learning algorithm [15]. It is a multilayer nonlinear feedforward network trained by an error back propagation learning algorithm. It consists of an input layer, a hidden layer, and an output layer. The BP learning algorithm consists of two processes: one is the forward propagation of data, and the other is the back propagation of error signals.

The interconnection mode between neural networks is formed by the interconnection of neurons, and the initial weight between each connection is randomly assigned by the computer. The forward propagation stage of data signal refers to the process of the original data signal from the input layer through the hidden layer to the output layer, that is, the output of the upper node is the input of the lower node. Each neuron y_i has a corresponding calculated weight w_{ik}, the output value o_{1k} of the input layer in the hidden layer is obtained by summing and weighting the input value, connection weight, and threshold t_k. The calculation method is as follows:

$$o_{1k} = \sum_{i=1}^{n} w_{ik} y_i + t_k. \tag{10}$$

In the process of prediction, better prediction accuracy can be obtained by using activation function processing. There are many kinds of activation functions, such as step function, ReLU functions, tanh functions, and Sigmoid functions [16]. In this paper, the Sigmoid function is used to activate the output information. The output value of the input layer is activated to $f(o_{1k})$ by the activation function, then the hidden layer z_k can be expressed as:

$$z_k = f(o_{1k}) = \frac{1}{1 + \exp(-o_{1k})}. \tag{11}$$

Next, the hidden layer data is used as the input layer to transfer the data to the output layer. The output value o_{2j} is weighted by the hidden layer value and the connection weight v_{kj} between the hidden layer, and the output layer, and then added to the threshold t_j. The calculation method is as follows:

$$o_{2j} = \sum_{k=1}^{n} v_{kj} z_k + t_j. \tag{12}$$

The output value is activated to obtain the data of the final output layer.

$$o_j = f(o_{2j}) = \frac{1}{1 + \exp(-o_{1j})}. \tag{13}$$

When the signal is transmitted to the output layer, the error function is used to detect whether the training process of the neural network is over. The stopping condition of the neural network is that it meets the limit value of the error function or reaches the set maximum number of iterations. When the output error function is less than the predetermined value, the training will stop. If the conditions are not met, the error will be back propagated.

The error function (Err) is used to measure the error between the actual output d_j and the expected output o_j.

$$\mathrm{Err} = \frac{1}{2}(d - o)^2 = \frac{1}{2}\sum_{j=1}^{n}\left(d_j - o_j\right)^2. \tag{14}$$

The error signal obtained from each layer is used to adjust the weight of the connection between neurons.

The process of back propagation is defined as follows:

$$\mathrm{Err} = \frac{1}{2}\sum_{j=1}^{n}\left(d_j - o_j\right)^2 = \frac{1}{2}\sum_{j=1}^{n}\left(d_j - f\left(\sum_{k=1}^{n} v_{kj} z_k\right)\right)^2. \tag{15}$$

By continuously adjusting the connection weight and threshold, the error is reduced along the gradient direction. After calculating the change value Δv_{jk} of the weight connection value between the hidden layer and the output layer, each connection weight is updated as follows:

$$\begin{aligned} \Delta v_{jk} &= -\beta \frac{\partial \mathrm{Err}}{\partial v_{jk}} = -\beta \frac{\partial \mathrm{Err}}{\partial o_{2j}} \frac{\partial o_{2j}}{\partial v_{jk}} = -\beta \frac{\partial \mathrm{Err}}{\partial o_{2j}} z_k, \\ v_{jk} &= v_{jk} + \Delta v_{jk}. \end{aligned} \tag{16}$$

The error is further extended to the input layer

$$Err = \frac{1}{2}\sum_{j=1}^{n}\left(d_j - f\left(\sum_{k=1}^{n} v_{kj} z_k\right)\right)^2 = \frac{1}{2}\sum_{j=1}^{n}\left(d_j - f\left(\sum_{k=1}^{n} v_{kj} f\left(\sum_{i=1}^{n} w_{ik} y_i\right)\right)\right)^2. \tag{17}$$

The connection weight between the input layer and the hidden layer is updated as follows:

$$\begin{aligned} \Delta w_{ik} &= -\beta \frac{\partial \mathrm{Err}}{\partial w_{ik}} = -\beta \frac{\partial \mathrm{Err}}{\partial o_{1k}} \frac{\partial o_{1k}}{\partial w_{ik}} = -\beta \frac{\partial \mathrm{Err}}{\partial o_{1k}} y_i, \\ w_{ik} &= w_{ik} + \Delta w_{ik}. \end{aligned} \tag{18}$$

This paper selects the students' physical fitness test data to establish the model, in which 80% of the students' samples are used as the training set and the other 20% are used as the test set for model evaluation. The number of neurons in the input layer is 8, the number of neurons in the output layer is 1, and the number of neurons in the hidden layer is 11. The threshold value is used as the condition value for the training stop. Its value is set to 0.005, and the learning rate is set to 0.1.

80% of the data after principal component analysis is brought into the function to establish the neural network model. In order to increase the accuracy of the model, the prediction models for boys and girls are established, respectively, and the accuracy of the model is evaluated by using the data of the test set.

In the process of experimental simulation, the mean square error (MSE) is used to compare the accuracy of the model prediction. In a comparison between the predicted value and the actual value of a sample of 40 students randomly selected in the test set, the mean square error of the model is 1.3612. It shows that the physical fitness test prediction model established in this paper is very accurate. The histogram of absolute error frequency distribution is shown in Figure 1.

As shown in Figure 1, the absolute error is between [−1, 1], which accounts for about 2/3 of the test data set, and the maximum error frequency is near 0. When the absolute error value is greater than 3, the error frequency is

Table 2: The transaction database D.

Tid	Items
T_1	I_1, I_2, I_4, I_5
T_2	I_1, I_2
T_3	I_2, I_4, I_5
T_4	I_2, I_3, I_4, I_5
T_5	I_1, I_4
T_6	I_1, I_2, I_5
T_7	I_1, I_3
T_8	I_1, I_2, I_3, I_4, I_5

Table 3: The calculation results of association rule confidence.

Item set association	Confidence
$I_2 \cap I_4 \longrightarrow I_5$	1.0
$I_2 \cap I_5 \longrightarrow I_4$	0.8
$I_4 \cap I_5 \longrightarrow I_2$	1.0
$I_2 \longrightarrow I_4 \cap I_5$	0.6
$I_4 \longrightarrow I_2 \cap I_5$	0.8
$I_5 \longrightarrow I_2 \cap I_4$	0.8

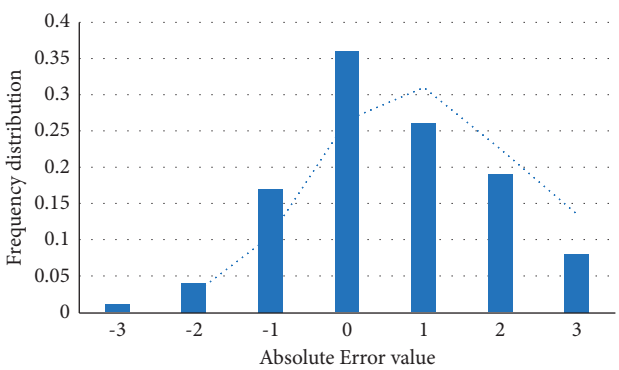

Figure 1: Error value frequency distribution.

very small, lower than 0.01. The dotted line is the error distribution density curve, and the error distribution of the model is similar to the normal distribution. The visualization results show that the prediction performance of the proposed model is very high.

5. Conclusion

In this paper, the physical test data and physical health self-assessment data of college students are taken as the research object. The decision tree C4.5 algorithm and association rule Apriori algorithm are selected to mine physical fitness data information. The classification law of students' physical health and the potential correlation between physical health factors are obtained. The related frequent itemsets are calculated according to the actual data, and the internal relationship between each index is obtained according to the support and confidence. On this basis, the comprehensive performance prediction model of college students' physical fitness test is constructed. The comprehensive performance prediction model of physical fitness tests successfully applies machine learning algorithms such as BP neural network and principal component analysis to the prediction of the comprehensive performance of physical fitness tests, and achieves more accurate prediction results. Physical fitness data types and collection forms are diverse. How to make an effective comprehensive analysis of different data resources is the further research direction in the future.

Data Availability

The basic data used in this paper is downloaded from the online public data set: Microdata of college students' physical health https://bbs.pinggu.org/a-3280384.html.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by a grant from the Shandong Province Soft Science Project of China (No. 2019RKB01267).

References

[1] R. S. Baker, "Educational data mining: an advance for intelligent systems in education," *IEEE Intelligent Systems*, vol. 29, no. 3, pp. 78–82, 2014.

[2] A. Kangaiammal, R. Silambannan, C. Senthamarai, and M. V. Srinath, "Student learning ability assessment using rough set and data mining approaches," *International Journal of Modern Education and Computer Science*, vol. 5, no. 5, pp. 1–11, 2013.

[3] N. Fang and L. U. Jingui, "A decision tree approach to predictive modeling of student performance in engineering dynamics," *International Journal of Engineering Education*, vol. 26, no. 1, pp. 87–95, 2010.

[4] A. Yadav, V. Alexander, and S. Mehta, "Case-based instruction in undergraduate engineering: does student confidence predict learning?" *International Journal of Engineering Education*, vol. 35, no. 1, pp. 25–34, 2019.

[5] R. K. Chunduri and A. K. Cherukuri, "Scalable algorithm for generation of attribute implication base using FP-growth and spark," *Soft Computing*, vol. 25, no. 14, pp. 9219–9240, 2021.

[6] J. Y. Bishop, H. M. Awan, D. M. Rowley, and R. W. Nagel, "Development and validation of a musculoskeletal physical examination decision-making test for medical students," *Journal of Surgical Education*, vol. 70, no. 4, pp. 451–460, 2013.

[7] L. Ramanathan, S. Dhanda, and D. S. Kumar, "Predicting students' performance using modified ID3 algorithm," *International Journal of Engineering and Technology*, vol. 5, no. 3, pp. 2491–2497, 2013.

[8] C. P. Maldonado-Erazo, A. Durán-Sánchez, J. Álvarez-García, and M. d. l. C. Del Río-Rama, "Sports sponsorship: scientific coverage in academic journals," *Journal of Entrepreneurship and Public Policy*, vol. 8, no. 1, pp. 163–186, 2019.

[9] W. Xu, W. Xiong, Z. Shao, and Y. Li, "Analysis of effectiveness and performance prediction of sports flipped classroom teaching based on neural networks," *Scientific Programming*, vol. 2021, Article ID 5284457, 7 pages, 2021.

[10] Y. Jin and Xiao-qiu, "Study on safety mode of dragon boat sports physical fitness training based on machine learning," *Safety Science*, vol. 120, no. 3, pp. 1–5, 2019.

[11] W. Li, X. Ma, Y. Chen et al., "Random fuzzy granular decision tree," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5578682, 17 pages, 2021.

[12] A. P. Muniyandi, R. Rajeswari, and R. Rajaram, "Network anomaly detection by cascading K-means clustering and C4.5 decision tree algorithm," *Procedia Engineering*, vol. 30, pp. 174–182, 2012.

[13] R. Zhang, W. Chen, T.-C. Hsu, H. Yang, and Y.-C. Chung, "ANG: a combination of Apriori and graph computing techniques for frequent itemsets mining," *The Journal of Supercomputing*, vol. 75, no. 2, pp. 646–661, 2019.

[14] B. Wang, D. Chen, and B. Shi, "Comprehensive association rules mining of health examination data with an extended FP-growth method," *Mobile Networks and Applications*, vol. 22, no. 2, pp. 1–8, 2017.

[15] D. S. Watson and M. N. Wright, "Testing conditional independence in supervised learning algorithms," *Machine Learning*, vol. 110, no. 8, pp. 2107–2129, 2021.

[16] J. Park and I. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, no. 2, pp. 246–257, 2014.