



## BLOCKCHAIN ENHANCED STUDENT PHYSICAL PERFORMANCE ANALYSIS USING MACHINE LEARNING-IOT AND APRIORI ALGORITHM IN PHYSICAL EDUCATION NETWORK TEACHING

JIANING LI<sup>\*</sup> ZHEPING QUAN<sup>†</sup> AND WEIJIA SONG<sup>‡</sup>

**Abstract.** In the digital era, particularly with the rise of online teaching, traditional approaches to college physical education face challenges in adequately monitoring and enhancing students' physical fitness. This study introduces a novel approach that integrates blockchain technology with a Machine Learning-IoT framework to evaluate and improve students' physical performance. Utilizing the Apriori algorithm, enhanced with particle swarm optimization and an improved K-means methodology, this system offers a robust tool for correlating student behavior with sports performance in a secure and decentralized manner. The proposed system uses blockchain for safe data management and IoT for real-time data collection, ensuring privacy as well as efficiency. The algorithm's accuracy, recall, and F1 values on the Iris dataset are 0.947, 0.931, and 0.928, respectively, with a considerable Calinski Harabasz score of more than 240. When applied to university student behavior data, the blockchain-enhanced system successfully mined association rules with a maximum confidence level of 0.923.

**Key words:** Apriori; Relevance; Clustering Algorithm; Network Teaching; Behavior Analysis; College Student

**1. Introduction.** The integration of sophisticated technologies such as the Internet of Things (IoT), Machine Learning (ML), and Blockchain has changed data analysis and decision-making processes in today's educational landscape. This is especially true in physical education, where the requirement to adequately evaluate and improve students' physical performance has become increasingly important. The introduction of online education practices has heightened the need for innovative ways that go beyond traditional limits.

With its interconnected network of physical devices, the Internet of Things provides an invaluable platform for real-time data collecting in educational contexts. This Internet of Things-based data gathering is critical for monitoring student actions and behaviors, especially in physical education, where performance indicators are dynamic and multidimensional. Machine Learning, on the other hand, provides powerful analytical capabilities for identifying patterns and insights in the massive amounts of data created by IoT devices. ML algorithms can help educators acquire a better understanding of student performance and develop more successful teaching tactics. By ensuring data integrity, security, and privacy, blockchain technology adds a key component to this ecosystem. Blockchain provides a decentralized and tamper-proof platform in educational situations where sensitive student data is involved, ensuring that the data utilized for analysis is reliable and safeguarded against unwanted access and manipulation.

In the context of educational reform in the new era, current college education emphasizes comprehensive development, in which physical education is an indispensable link. With the gradual improvement of information technology construction, the teaching of various subjects in universities is gradually networked [1]. Against this background, the current problem of less attention paid to physical education and the cultivation of students' physical fitness in college education is even more obvious [2]. Therefore, it is necessary to provide a practical evaluation tool for online teaching of physical education in universities. Apriori is the most classical and widely used association rule mining algorithm [3]. The algorithm uses a layer by layer iterative search method to mine hidden Boolean association rules between data [4]. K-means is one of the main tools for data analysis, which performs tasks such as data classification through clustering analysis [5]. This study introduces K-means into Apriori algorithm to construct a correlation algorithm between student behavior and sports performance to

---

<sup>\*</sup>College of Physical Education, Taiyuan Normal University, Jinzhong, 030619, China ([facejob2023@163.com](mailto:facejob2023@163.com))

<sup>†</sup>College of Physical Education, Taiyuan Normal University, Jinzhong, 030619, China (Corresponding author, [zhepingquan@outlook.com](mailto:zhepingquan@outlook.com))

<sup>‡</sup>Wesleyan College, Cavite State University, Manila, 0900, The Philippines ([songweijia199472@163.com](mailto:songweijia199472@163.com))

optimize the current situation of online sports teaching and help improve students' physical fitness. During the construction process, the K-means clustering results and the efficiency shortcomings of Apriori are also optimized. The purpose of this study is to bring practical results to the field of online physical education in universities. This research presents a comprehensive solution that synergizes IoT, ML, and Blockchain technologies to analyze and enhance student physical performance in an educational context. The main contributions of this research are:

1. We present a novel architecture that blends IoT's real-time data collection capabilities with ML's analytical prowess, all while being supported by blockchain technology's security and integrity. This comprehensive technique ensures a comprehensive examination of student physical performance while protecting data privacy and security.
2. We have adapted the Apriori algorithm, which has historically been employed in market basket analysis, for use in physical education. This modification, enhanced with particle swarm optimization and better K-means techniques, provides an effective tool for linking student behavior with sports success, giving instructors with actionable data.
3. Our research addresses the crucial concerns of data integrity and privacy in educational data analysis by utilizing blockchain technology. This ensures that the findings of the study are founded on secure and accurate data, which is critical in educational settings.

**2. Related works.** Apriori, as a major algorithm in the field of association mining, has been widely studied and applied so far. Hazelton J proposed a cost estimation model for engineering calculation based on Apriori association mining. Experimental results showed that the model could accurately estimate the cost of engineering construction [6]. Karthik S and Velu CM proposed a peak season sales forecasting system based on user data sets using Apriori algorithm instead of traditional algorithms [7].

According to the comparison results between this system and traditional algorithms, the accuracy, sensitivity, and specificity of this system reached 73%, 78%, and 80%, respectively, which were much higher than traditional algorithms. Moreover, the running time of this system was also shorter, so it had higher applicability. Sornalakshmi M et al. provided the healthcare industry with a new improved Apriori algorithm that can separate frequent itemsets and delete abnormal data at the beginning of the algorithm, thereby reducing the resource requirements for algorithm operation [8]. According to the test results, the results generated by this algorithm in the local minimum support degree of healthcare databases were relatively reliable, and can effectively reduce medical costs. K-means, as a widely used clustering algorithm, also had a lot of related research. Sun Z et al. proposed a short-term traffic flow prediction model based on K-means clustering and gated recursive units. This model can predict the traffic flow in the future based on the characteristics of historical traffic flow data [9]. Compared with methods such as random forest and support vector machine regression, this method fully considered the pattern diversity of traffic flow, and had higher prediction accuracy. Kumar R U and Jeeva J B proposed a color cutting method using the clustering analysis function of K-means, which was used to evaluate and predict the healing of skin surface wounds [10]. This method was non-invasive and can effectively reduce the pain of patients, thereby reducing the tension and anxiety of patients facing medical facilities. Nyanjara S et al. with the help of experts to develop an assessment model for maternal and neonatal health quality in developing countries, which is based on K-means [11]. The model can classify and summarize the health data of the target population, and assign different data points to the most appropriate clustering. The test results showed that the classification accuracy of this method exceeded 73%. In the field of education, the teaching of physical education courses has always been one of the focuses of attention of educators. Huang Y Y studied the integration of online courses and physical education in current higher vocational education, and provided guiding suggestions for online physical education [12]. Guo J and Sun C proposed a real-time monitoring system for physical education classes based on the Internet of Things and cloud computing, which can automatically locate students' positions in sports scenes and evaluate the quality of their courses [13]. The system to some extent achieved the monitoring of students' performance in physical education courses, but failed to comprehensively evaluate students' physical exercise both inside and outside of class. Wang GR analyzed the current situation of physical education teaching in schools from a medical perspective [14]. According to the analysis conclusion, the current physical education teaching industry lacked teachers with professional medical knowledge, and the physical education teaching in various schools also lacked medical characteristics.

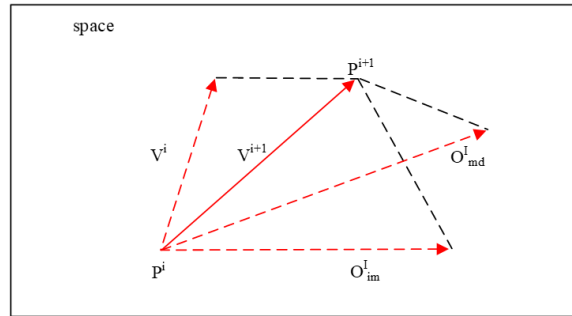


Fig. 3.1: Theory of Particle Swarm Optimization

The study believed that physical education combined with medical theory can reduce accidental injuries among students and further increase their exercise effectiveness.

Through sorting out relevant research results, it is found that most of the research on physical education teaching in universities focused on improving the quality of courses and the remoteness of the network. However, students' physical performance exists not only in physical education classes, but also in various aspects outside of class. Therefore, a comprehensive evaluation tool is necessary. Although Apriori and K-means are widely used, there is still a gap in intelligent sports teaching. Therefore, this study proposes a correlation algorithm between student behavior and sports performance based on Apriori.

### 3. Apriori Algorithm Construction for Correlativity between Student Behavior and Sports Performance.

**3.1. Adaptive Weighted K-means Algorithm for Apriori Analysis.** The behavior data of students is mainly continuous data, such as consumption data, network access data, etc. Due to the small dimensions and values of such data, K-means algorithm is suitable for clustering analysis. This is because K-means is a machine learning which has high efficiency and scalability, and low implementation costs [15]. However, the output effect of K-means is highly dependent on the selection of cluster centers. The selection of cluster centers is random, and the number of cluster centers is set by the operator through experience and test results, with a low degree of automation. In addition, K-means also has problems with local optimal solutions and unstable clustering results [16]. To address these shortcomings and make K-means more efficient in serving the processing of student behavior data, an optimized K-means based on Particle Swarm Optimization (PSO) is proposed which resembles like deep learning model. The principle of PSO is shown in Figure 3.1. In multi-dimensional space, each particle is searching for other individuals, and they constantly obtain the location of the next search, and continue to iterate until they find the optimal location. Introducing PSO into K-means can effectively optimize the randomness of cluster center selection, ensuring that K-means can find relatively excellent initial cluster centers. In addition, an adaptive weighting strategy is proposed to solve the problem of unstable clustering results.

Firstly, the PSO optimization part is constructed. In a search space with dimension  $M$ , the number of particles is  $N$ . The position of particle  $i$  can be expressed by equation 3.1.

$$p_{1m}, p_{2m}, \dots, p_{im} \tag{3.1}$$

The  $P_{im}$  in equation 3.1 represents the position of particle  $i$ . The velocity of the particle is set to  $V$ , and the velocity vector of the particle is shown in equation 3.2.

$$V_{im} = (v_{i1}, v_{i2}, \dots, v_{im}) \tag{3.2}$$

The position and velocity of particle is  $i$ . The optimal solution of its individual is  $V$ , as shown in equation 3.3.

$$\begin{cases} O_{im} &= (o_{i1}, o_{i2}, \dots, o_{im}) \\ O_d &= (o_{1,d}, o_{2,d}, \dots, o_{M,d}) \end{cases} \quad (3.3)$$

In equation 3.3,  $O_d$  represents the group optimal solution. After obtaining individual and group optimal solutions, the particle swarm can update its iteration speed and position. The mathematical expression of the velocity vector  $V_{im}^k$  of the updated particle in the  $m$ th dimension is shown in equation 3.4.

$$V_{im}^I = \omega v_{im}^{I-1} + (o_{im}^I - p_{im}^I)c_1 r_1 + (O_{md}^I - p_{im}^I)c_2 r_2 \quad (3.4)$$

In equation 3.4,  $I$  represents the current number of iterations.  $\omega$  is the inertia weight, and the weight value is proportional to the global optimization ability of the algorithm, while the local optimization ability is inversely proportional.  $c_1$  and  $c_2$  represent learning factors, the former being individual learning factors, and the latter being group learning factors.  $r_1$  and  $r_2$  are the random number used to increase the randomness of the search, and the values of both are  $[0,1]$ .  $o_{im}^I$  and  $O_{md}^I$  respectively identify the historical optimal positions of individuals and groups. The position vector  $P_{im}^I$  of the particle in  $m$ th dimension after updating is shown in equation 3.5.

$$p_{im}^I = p_{im}^{(I-1)} + v_{im}^{(I-1)} \quad (3.5)$$

After completing the construction of the particle swarm, you can introduce it into K-means. It is necessary to first calculate the particle fitness to obtain the global optimal solution. After inputting the optimal solution information and the data to be predicted, the reciprocal of the total distance is calculated from all sample points to the corresponding cluster center, as shown in equation 3.6.

$$\text{FIT} = \sum_{k=1}^K \sum_{p_i \in P_K} d(p_i, C_k) \quad (3.6)$$

In equation 3.6, FIT represents the fitness function value, and  $d(p_i, C_k)$  is the distance from the sample point to its corresponding cluster center.  $K$  refers to the k-value of the algorithm. The role of particle swarm optimization in this algorithm is to optimize the selection of clustering centers. In addition, adaptive weights should be used to optimize the performance instability of K-means. Assuming that there are  $n$  data to be processed with a dimension of  $m$ , the data can be converted into a matrix as shown in equation 3.7.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (3.7)$$

If the number of clusters is the same as the K value of the algorithm, the total intra class distance of all clusters in the algorithm on the  $u$ th dimension is shown in equation 3.8.

$$D_{in} = \sum_{k=1}^K \sum_{j=1}^n (x_{ju} - m_{ku})^2 \quad (3.8)$$

In equation 3.8,  $m_{ku}$  is the mean value of the k-th cluster on the attributes of the  $u$ th dimension. After obtaining this value, you can further obtain the sum of the inter class distances of all clusters on this dimension attribute, as shown in equation 3.9.

$$D_{out} = \sum_{k=1}^K (m_{ku} - m_u)^2 \quad (3.9)$$

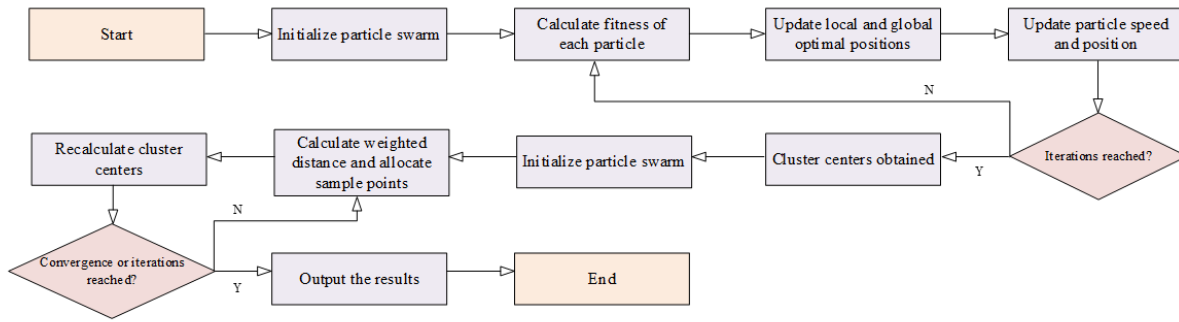


Fig. 3.2: The Improved K-means Algorithm for Student Behavior Data Processing

From this, it is possible to calculate the impact of the attributes of the  $u$ th dimension on the clustering results, as shown in equation 3.10.

$$\text{impact}_u = \frac{D_{in}}{D_{out}} \tag{3.10}$$

When a sample in a space has compact intra cluster distances and large inter cluster distances, it indicates that the clustering results are good, further indicating that the attribute dimension has a significant impact on the clustering results. Assigning weights to attributes with different dimensions and giving greater weights to more important attributes can optimize the clustering results of the entire algorithm. Set the attribute weight of dimension  $u$  to  $W$ , as shown in equation 3.11.

$$W_u = \frac{\text{impact}_u}{\sum_{u=1}^M \text{impact}_u} \tag{3.11}$$

In equation 3.11, the value of each  $W$  is between 0 and 1, and the sum of  $W$  for each dimension is 1. Adding the calculated attribute weights to the traditional K-means Euclidean distance equation can obtain a new distance equation, as shown in equation 3.12.

$$D(a, b) = \sqrt{\sum_{u=1}^M W_u (x_{au} - x_{bu})^2} \tag{3.12}$$

The flow of the improved K-means algorithm for student behavior data processing is shown in Figure 3.2. Before performing the K-means operation, the number of cluster centers is first obtained through PSO, and then the weights are initialized. In this algorithm, sample points are calculated based on weighted Euclidean distances, and then K-means iteration is performed. Compared with traditional K-means, this algorithm solves the problems of cluster center selection and sample calculation stability through PSO and weighted distance.

Apriori is one of the most important algorithms in the field of data mining, and its specific process is shown in Figure 3.3. Based on the preset support and confidence levels, the algorithm will determine the threshold value of the candidate item set and decide to prune or access the next step [17]. When the processed data is greater than or equal to the minimum confidence and support, it is considered that there is correlation between the data, and the algorithm can derive its association rules [18]. However, Apriori running according to this process has two main drawbacks. Firstly, the algorithm will generate a large number of candidate itemsets, a large portion of which are unnecessary for computation, which has a negative impact on the efficiency of the algorithm and increases the computational cost [19]. Secondly, Apriori scans the transaction database too much during the calculation process, which leads to a significant increase in the algorithm’s calculation time, seriously affecting the efficiency of the algorithm [20].

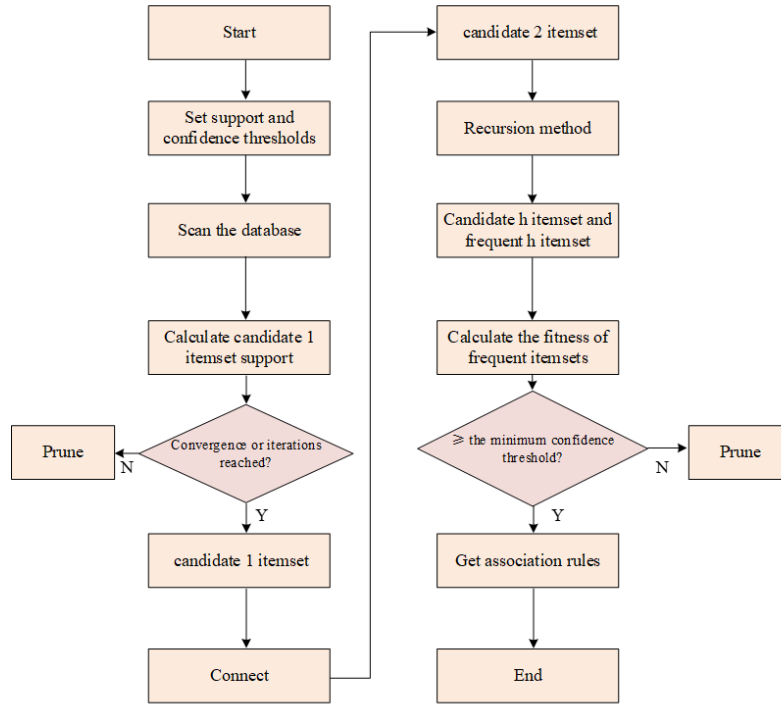


Fig. 3.3: Structure of traditional Apriori algorithm

The analysis of student behavior data usually involves the entire school’s student data, and the amount of data is relatively large, so there are high requirements for algorithm efficiency. To improve the efficiency of Apriori, an optimization method based on pre pruning is proposed to address its main shortcomings. Pruning is an operation of Apriori to filter itemsets that do not meet the support and confidence levels, and must be run at least twice in a round of algorithm runs. Therefore, performing pre pruning during the operation of the algorithm can reduce the number of candidate itemsets and the number of times the algorithm scans the transaction database, effectively improving the two main defects of Apriori. There is a set of frequent items, and data appears times in its subset, equation 3.13 can be obtained.

$$|L_{h-1}(z)| \geq h - 1 \tag{3.13}$$

In equation 3.13,  $L_{h-1}(z)$  refers to the number of occurrences of in frequent itemset  $L_{h-1}$ . According to this equation, the conditions for judging the  $h$  term set  $Y$  as an infrequent term set are shown in equation 3.14.

$$\begin{cases} |L_{h-1}(z)| < h - 1 \\ z \in Y \end{cases} \tag{3.14}$$

When  $Y$  is a non frequent itemset, determine whether the  $h$  itemset connected to the  $Y_j$  existing in itemset  $Y$  is a non frequent itemset, and the judgment conditions are shown in equation 3.15.

$$\begin{cases} |L_{h-1}(z)| < h - 1 \\ \exists z \in Y_j \end{cases} \tag{3.15}$$

When equation 3.15 is satisfied,  $Y_j$  does not need to add connections, and it is pruned during the operation of the algorithm, thereby increasing the efficiency of the algorithm. The improved Apriori algorithm flow is

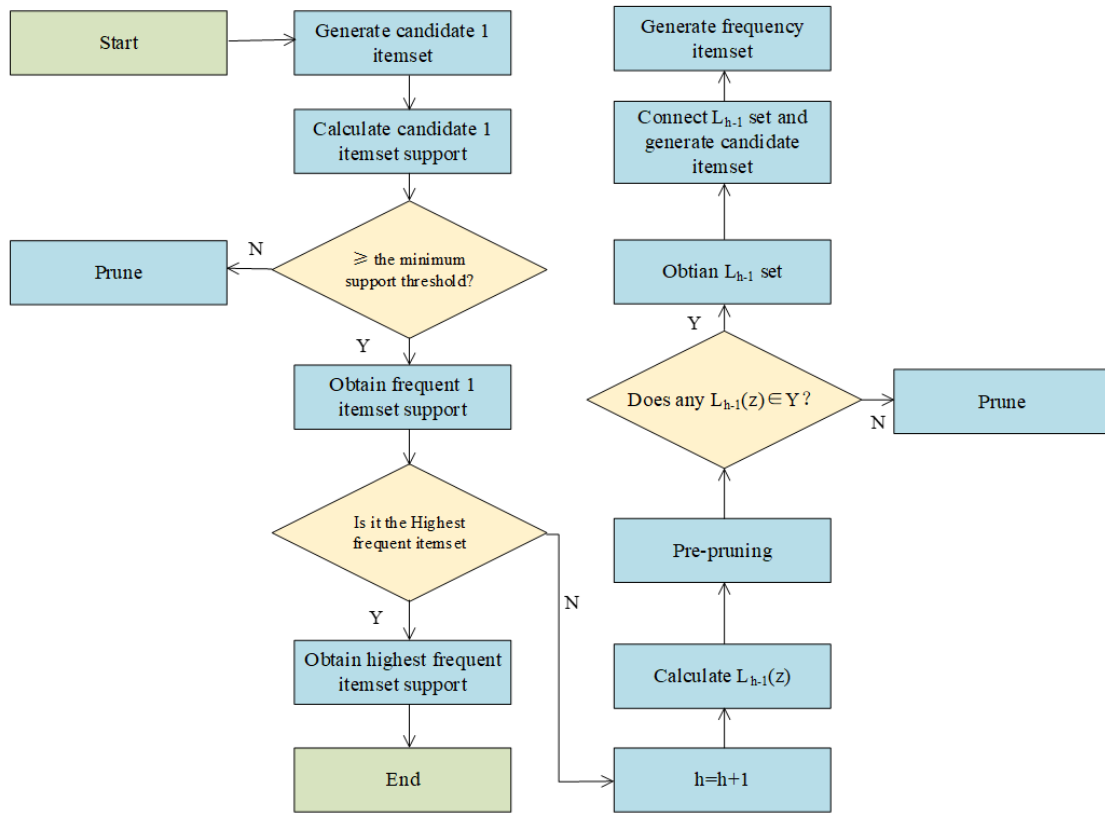


Fig. 3.4: Improved Apriori Operation Flow

shown in Figure 3.4. The data processed by improved K-means clustering is input into the algorithm and a candidate item set is generated. During the operation of the algorithm, some steps, including calculating confidence levels, are replaced with pre pruning steps, thereby reducing the amount of itemset generation and scanning times, and increasing operational efficiency.

After completing the construction of the correlation algorithm, specific settings need to be made for student data processing and correlation analysis. First, the characteristics and corresponding labels of different types of students are designed based on the type of student behavior data, as shown in Table 3.1. Student behavior is divided into three categories, including consumer behavior, life behavior, and sports behavior. The information included in consumption behavior mainly includes total sales volume, single consumption fluctuation, consumption volume for three meals, consumption frequency, and additional sales volume not necessary for daily life. Consumer behavior can reflect students' living habits and daily activities to a certain extent, and is an important evaluation dimension. The information included in life behavior mainly includes the time of students' meals, time of getting up and sleeping, length of sleeping, and time of surfing the internet. Life information reflects whether students' lives are regular, healthy, and have time for exercise. Healthy living habits are the foundation of excellent sports performance and physical fitness, so this is a major evaluation dimension. The classification of students' living habits is mainly based on whether their life data are regular. The last type of behavior is sports behavior, which mainly includes information about the number and time of entering the stadium and gymnasium, as well as the length of each exercise, and the use of sports equipment. This dimension of information can most directly reflect the students' sports situation.

After defining the behavioral characteristics and corresponding labels of students, it is also necessary to define their physical performance. The definition of physical education performance is based on the students'

Table 3.1: Indicator Labels for Association Rule

Categories	Features	Tags
Consumer behavior	Low living expenses, stable and regular consumption, basically no extra expenses	Low expenditure type
	The cost is too high, the single consumption fluctuates greatly, the takeout consumption frequency is high, and the extra cost is high	High expenditure type
	Low living expenses, stable and regular consumption, low extra expenses	Frugal type
	Moderate spending, moderate consumption frequency and regular consumption	Moderate type
	Low consumption frequency, low cost of meals, and high extra cost	In campus low expenditure type
Daily life behavior	Irregular meals, dependence on takeout, basically not getting up early, long internet time	Irregular type
	Regular meals, occasional takeout, frequent early getting up, short internet time	very regular type
	Meals are irregular, the number of takeouts is small, the early getting up is not frequent, and the internet time is moderate	regular type
	Irregular meals, seldom get up early, often order takeout, and spend a long time online	less regular type
Sports behavior	The equipment is highly used, the number of times of admission is many, and the exercise time is long	Very effort type
	Average use of equipment, many times of admission and long exercise time	effort type
	The use of equipment is small, the number of admission is average, and the exercise time is short	less effort type
	Less equipment use, less admission times and less exercise time	effortless type

annual physical examination performance and the results of each semester’s physical education courses. Sports performance is classified using a comprehensive scoring system. The score is calculated based on the weighted average score on which the definition of physical performance is based. The scoring range is 1 to 5 points, with 4-5 points representing outstanding physical performance, 3-4 points representing less outstanding physical performance, 2-3 points representing average physical performance, and 1-2 points representing Bad physical performance. When a student’s physical performance is 1-2 points, their physical performance is extremely poor and they fail in terms of physical fitness.

**4. Apriori Algorithm Experiment and Application for Correlating Student Behavior and Sports Performance.**

**4.1. Correlation Algorithm Performance Test.** To evaluate the value of the proposed algorithm, it is necessary to test the algorithm. The test consisted of two parts, namely, algorithm performance test and practical application test. Performance testing tested the computational capabilities and characteristics of the improved K-means and Apriori algorithms in the algorithm. Due to the significant impact of hardware and software environments on algorithm performance, this experiment was conducted in a fixed environment configuration. The detailed configuration is shown in Table 4.1. The experimental operating system was Windows 10, and the programming environment was Python. The data used for the test was from the UCI database, where three datasets, Iris, Glass, and Wine, were used to test the performance of the algorithm.

Firstly, the clustering accuracy of the proposed algorithm was tested. Because the clustering of the algo-



Table 4.1: Configuration Related to Performance Test

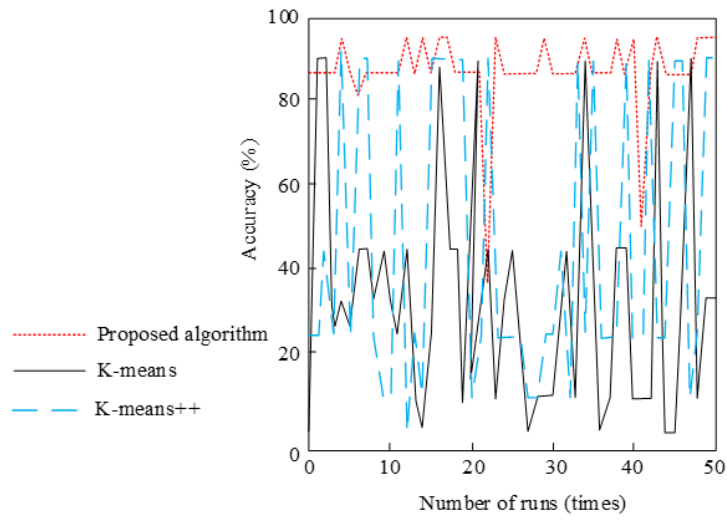
CPU	Intel(R) i7-8700k
RAM	16GB
Operating system	Windows 10
Hard disk	1 TB HDD
Programming environment	Python
Experimental database	UCI

Table 4.2: Assessment Data of the Proposed Algorithm and Comparative Algorithms

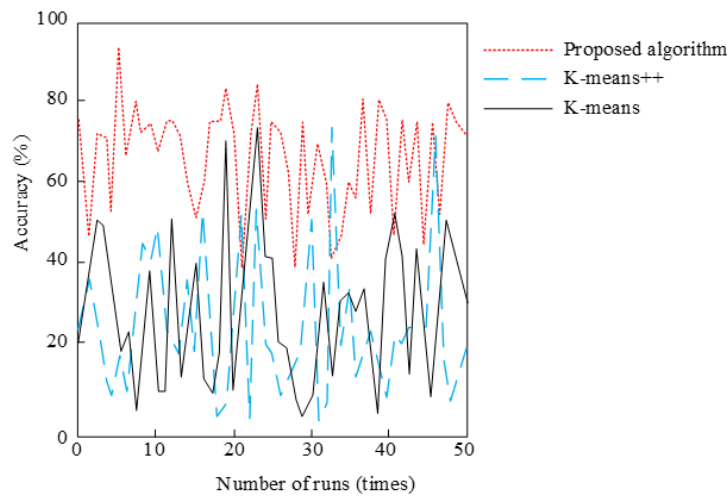
Data sets	Algorithm	Precision	Recall	F1
Iris	K-means	0.907	0.874	0.892
	K-means++	0.907	0.893	0.895
	Proposed algorithm	0.947	0.931	0.928
Wine	K-means	0.531	0.455	0.470
	K-means++	0.548	0.443	0.481
	Proposed algorithm	0.594	0.497	0.509
Glass	K-means	0.135	0.197	0.127
	K-means++	0.138	0.158	0.159
	Proposed algorithm	0.342	0.313	0.233

rithm was performed by an improved K-means algorithm, the traditional K-means algorithm and the improved K-means++ algorithm based on K-means were used here as a comparison algorithm. The results of the accuracy test are shown in Figure 4.1. Iris and Wine datasets were used to test the accuracy of centralized algorithms under different data sets. The three algorithms were run 50 times on each dataset. Figure 4.1a shows the performance of the three algorithms in the Iris dataset, and Figure 4.1b shows the performance of the three algorithms in the Wine dataset. Under the two datasets, the accuracy rates of the three algorithms showed significant fluctuations, which was caused by the characteristics of the K-means algorithm itself. Observing the graph, the proposed algorithm exhibited minimal fluctuations in both data sets, and its accuracy curve was mostly higher than K-means algorithm and K-means++ algorithm. In the Iris dataset, the accuracy rate of the proposed algorithm was above 80% for most of the time, and both comparison algorithms were below this level. Precision, recall, and F1 values are also important indicators for evaluating algorithm performance. After measuring the average data of 50 runs of the three algorithms, their above indicators were compared, as shown in Table 4.2. The performance of K-means algorithm and K-means++ were relatively close, and the two algorithms had advantages and disadvantages for each other under different data sets. The precision, recall, and F1 of the proposed algorithm were steadily higher than those of the other two algorithms under the three data sets. Its average precision, recall, and F1 values under the Iris dataset were 0.947, 0.931, and 0.928, respectively. In the same dataset, the recall rate and F1 values of K-means were 0.874 and 0.892, respectively, for K-means++, which are 0.893 and 0.895. Combining the accuracy test results, the proposed algorithm outperformed K-means and K-means++ algorithms in clustering ability, which laid a solid data foundation for later correlation analysis.

For cluster analysis, the experiment also used its unique Silhouette Coefficients (SC) and Calinski Harabasz Score (CH) for evaluation. SC showed the differences within and outside the K-means cluster, and larger differences represented better clustering results. CH is an indicator obtained by calculating the ratio of inter cluster variance to intra cluster variance under certain conditions. The higher the value of this indicator, the better the clustering effect. The results of SC and CH evaluations are shown in Figure 4.2. Figure 4.2a shows the SC test results of the algorithm under different data sets, and Figure 4.2b shows the CH test results of the algorithm under different data sets. There are significant differences in the performance of each algorithm within different data sets depending on the different data sets. Observing within the same dataset, the SC



(a) Accuracy Under Iris

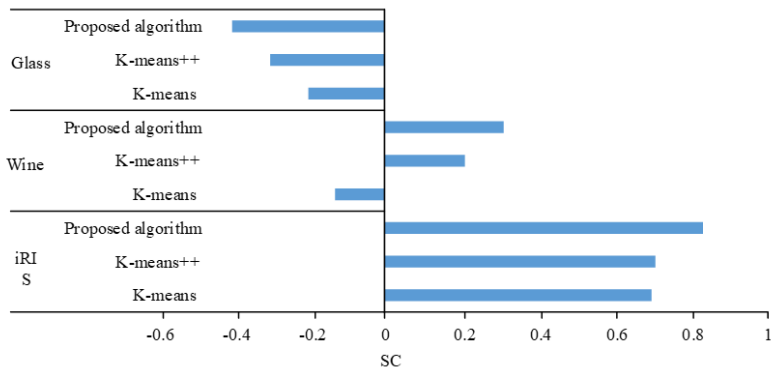


(b) Accuracy Under Wine

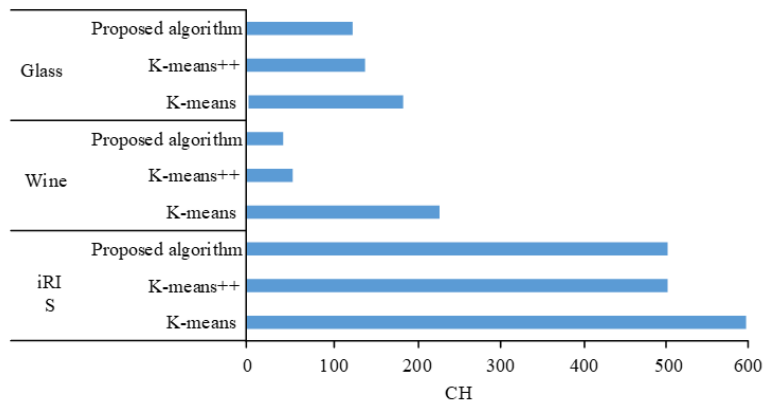
Fig. 4.1: Results of Accuracy Test under Different Data Set

and CH of the proposed algorithm were both higher than K-means and K-means++. Under the Wine dataset, the CH of the proposed algorithm was 240.3, while the CH of K-means and K-means++ were 63.4 and 67.2, respectively. Based on the evaluation results of SC and CH, the proposed algorithm had better clustering ability in general.

Due to the large amount of data on students' behavior data and sports performance data, the speed of algorithm operation was extremely important on the premise of ensuring correctness. The proposed algorithm mainly reduced the amount of item set processing and scanning times during Apriori's operation through pre pruning, thereby reducing the burden and cost of correlation calculation and increasing efficiency. To verify the



(a) SC Under Different Dataset



(b) CH Under Different Dataset

Fig. 4.2: SC and CH Evaluation Results of Different Algorithms

effectiveness of this optimization, the traditional Apriori, Dual Searching Apriori, and Pre Apriori proposed in this study were combined with K-means and the PSO-K-means proposed in this study. Their operational efficiency against the same dataset was tested. The results are shown in Figure 4.3. The curve of the proposed algorithm was below the other comparison algorithms, indicating that the proposed algorithm had the fastest running speed. When the support threshold was 0.4, the computation time was about 3400ms, significantly lower than other algorithms.

**4.2. Correlation Algorithm Practical Application Test.** After verifying the performance of the algorithm in the test dataset, it was also necessary to test it under a real student behavior dataset to verify its practical application value. A university was selected as the main venue for testing. The university has a high degree of intelligence in student management. Through the campus network management system, it was possible to query students' consumption on campus, the time to enter and leave the school and gym, the online duration of the campus network, dormitory water consumption, and light off time. This provided conditions for testing the correlation between student behavior and physical performance. After negotiation, the behavioral data and sports performance of 1540 students in a certain grade of this size were selected as data samples, and the samples were sent to the proposed algorithm for clustering analysis and mining related rules. According to the previous test results, the minimum support threshold of the algorithm was set to 0.2, and the minimum confidence threshold was set to 0.4. After mining, a total of 37 relevant rules were mined, as

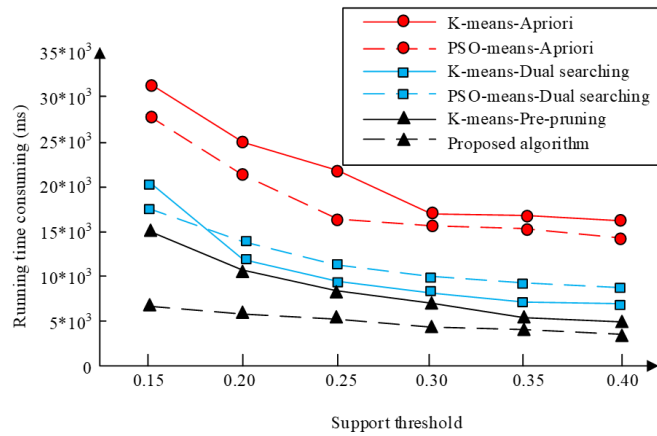


Fig. 4.3: Comparison Results of Running Time of Algorithms

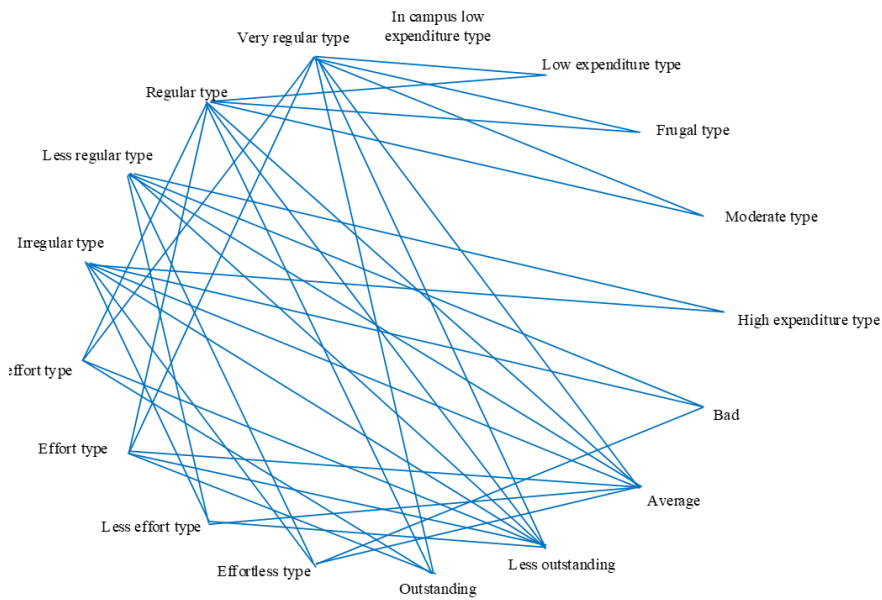


Fig. 4.4: Display of Association Rule Mining Results

shown in Figure 4.4. Through the images, there were more association rules between students' life behavior and sports behavior and their sports performance, while there were fewer association rules between consumer behavior and sports performance.

Table 4.3 presents the results of association rule mining in a data format, with support and confidence levels indicated after each rule. Two association rules treated sports as outstanding and less outstanding as post rules. Five association rules used average as a post rule for sports performance. Three association rules treated sports as bad as a post rule. Among the association rules mined, the highest support level was 0.550, and the lowest was 0.212. The highest confidence level was 0.923 and the lowest was 0.727. From the mining results of association rules, sports performance was related to life rules and sports habits. Students who had a regular, active lifestyle, and long-term exercise habits were more likely to have excellent physical performance.

Table 4.3: Association Rule Mining Results

Rule	Support	Confidence
Very regular type, very effort type: outstanding	0.451	0.862
Regular type, very effort type: outstanding	0.422	0.841
Very regular type, effort type: less outstanding	0.520	0.736
Regular type, effort type: less outstanding	0.513	0.727
Very regular type, less effort type: average	0.345	0.747
Less regular type, less effort type: average	0.424	0.912
Less regular type, less effort type, moderate type: average	0.235	0.785
Low expenditure type, less effort type: average	0.212	0.772
Frugal type, very regular type, less effort type: average	0.434	0.783
Effortless type: bad	0.546	0.923
Irregular type: bad	0.550	0.895
Effortless type, irregular type: bad	0.535	0.883

**5. Conclusion.** To improve the physical quality of contemporary college students and promote their comprehensive development, a mining algorithm based on Apriori for the association between college students' behavior data and sports performance is proposed to address the current lack of physical performance evaluation methods and effectiveness. The algorithm combines an improved K-means clustering and Apriori correlation algorithm. According to the experimental results, the average accuracy, recall, and F1 values of the proposed algorithm in the Iris dataset were 0.947, 0.931, and 0.928, respectively. The indicators of other algorithms in the same dataset were lower than those of the proposed algorithm. Under the Wine dataset, the CH of the proposed algorithm was 240.3, while the CH of K-means and K-means++ did not exceed 100. When the support threshold was 0.4, the computation time of the proposed algorithm was about 3400ms, significantly lower than other algorithms. In practical applications, the proposed algorithm had mined 12 association rules, with a maximum support level of 0.550 and a minimum confidence level of 0.212. The maximum confidence level was 0.923 and the minimum confidence level was 0.727. The results of association rule mining provided reference information based on quantitative data for the formulation of college physical training measures and the improvement of students' physical fitness. Although the proposed algorithm has already been practical, there is still room for improvement. The current construction of student behavior data system is not perfect. In further research, more important behavior data will be screened through controlled variable experiments, and less important behavior data will be eliminated.

**Funding.** The research is supported by The humanities and social science planning project of the Ministry of Education, "Research on the security mechanism of deepening the integration of sports and education to promote the implementation of the adolescent physical health policy", subject number: 21YJA890024..

#### REFERENCES

- [1] Qi, S., Li, S. & Zhang, J. Designing a Teaching Assistant System for Physical Education Using Web Technology[J]. *Mobile Information Systems* (. 6 pp. 1-11 (2021)
- [2] Wang, Z. Investigation on the Employment of Ex-Soldiers in Physical Education Major in Colleges and Universities [J]. *Open Access Library Journal*. 8, 1-8 (2021)

- [3] Zhang, Z., Liu, X., Li, Z. & Hu H. Outburst prediction and influencing factors analysis based on Boruta-Apriori and BO-SVM algorithms[J]. *Journal Of Intelligent And Fuzzy Systems*. **2021**, 1-18 (0)
- [4] Aria, R. & Susilowati, S. Analisa Data Penjualan SaRa Collection menggunakan metode Apriori[J]. *Jurnal Teknik Komputer*. **7**, 68-73 (2021)
- [5] Agalya, V., Kandasamy, M., Venugopal, E. & Others CPRO: Competitive Poor and Rich Optimizer-Enabled Deep Learning Model and Holoentropy Weighted-Power K-Means Clustering for Brain Tumor Classification Using MRI[J]. *International Journal Of Pattern Recognition And Artificial Intelligence*. **36**, 1-30 (2022)
- [6] J., H. THE AUTOMATING OF COST ESTIMATES: APRIORI AND ITS SOFTWARE[J]. *F & M: Fabricating & Metal-working*. **21**, 44-471 (2022)
- [7] Karthik, S. & Velu, C. Novel prediction of sales and purchase forecasting for festival season of Hypermarkets with customer dataset using Apriori algorithm instead of FP-Growth algorithm to improve the accuracy[J]. *ECS Transactions*. **107**, 12647-12659 (2022)
- [8] Sornalakshmi, M., Balamurali, S., Venkatesulu, M., Krishnan, M., Ramasamy, L., Kadry, S. & Lim, S. An efficient apriori algorithm for frequent pattern mining using mapreduce in healthcare data[J]. *Bulletin Of Electrical Engineering And Informatics*. **10**, 390-403 (2021)
- [9] Sun, Z., Hu, Y., Li, W., Feng, S. & Pei, L. Prediction model for short-term traffic flow based on a K-means-gated recurrent unit combination[J]. *IET Intelligent Transport Systems*. **16**, 675-690 (2022)
- [10] Kumar, R. & Jeeva, J. Segmentation of Wound By K-Means Clustering and Automatic Prediction of Healing Time[J]. *ECS Transactions*. **107**, 20371-20376 (2022)
- [11] Nyanjara, S., Machuve, D. & Nykanen P. Maternal and Child Health Care Quality Assessment An Improved Approach Using K-Means Clustering[J]. *Journal Of Data Analysis And Information Processing*. **10**, 170-183 (2022)
- [12] Huang, Y. Integrating Online Teaching into Public Physical Education - Taking Vocational Colleges in Chongqing as Examples[J]. *JOURNAL OF CONTEMPORARY EDUCATIONAL RESEARCH*. **6**, 102-107 (2022)
- [13] Guo, J. & Sun, C. Real-time monitoring of physical education classroom in colleges and universities based on open IoT and cloud computing[J]. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology*. **40**, 7397-7409 (2021)
- [14] Wang, G. Exploration of Teaching and Education: Integration of Sports and Medicine in Physical Education in Colleges and Universities[J]. *JOURNAL OF CONTEMPORARY EDUCATIONAL RESEARCH*. **6**, 63-67 (2022)
- [15] Warchalska-Troll, A. & Warchalski, T. The selection of areas for case study research in socio-economic geography with the application of k-means clustering[J]. *Wiadomosci Statystyczne*. **67**, 1-20 (2022)
- [16] Majdina, N., Soeleman, M. & Supriyanto, C. Application of Particle Swarm Optimization (PSO) to Improve K-means Accuracy in Clustering Eligible Province to Receive Fish Seed Assistance in Java[J]. *IOSR Journal Of Computer Engineering*. **24**, 43-49 (2022)
- [17] Qisman, M., Rosadi, R. & Abdullah, A. Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores) [J]. *Journal Of Physics: Conference Series*. **1722** pp. 1 (2021)
- [18] Javed, M., Nawaz, W. & Hova-fppm, K. Flexible Periodic Pattern Mining in Time Series Databases Using Hashed Occurrence Vectors and Apriori Approach[J]. *Scientific Programming* (. **1** pp. 1-14 (2021)
- [19] Muchlis, M., Fitri, I. & Nuraini, R. Rancang Bangun Aplikasi Data Mining pada Penjualan Distro Bloods Berbasis Web menggunakan Algoritma Apriori[J]. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*. **4**, 26-33 (2021)
- [20] Saha, S., Prasad, A., Chatterjee, P., Basu, S. & Nasipuri M. Modified FPred-Apriori improving function prediction of target proteins from essential neighbours by finding their association with relevant functional groups using Apriori algorithm[J]. *International Journal Of Advanced Intelligence Paradigms*. **19**, 61-83 (2021)

*Edited by:* Kumar Abhishek

*Special issue on:* Machine Learning and Block-chain based Solution for Privacy and Access Control in IoT

*Received:* Oct 7, 2023

*Accepted:* Dec 19, 2023