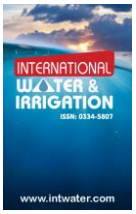


International Water and Irrigation
Volume 44 Issue 1 (2025)
ISSN: 0334-5807
Pages 140 - 146

Intelligent Analysis of College Physical Education Teaching Effect Based on Data Mining Technology

Yuan Zhang[1][0000-1111-2222-3333]

[1]Sports Department of Hubei University of Automotive Technology, Shiyan, Hubei, 442002, China
happyearthren@qq.com

Abstract. This study enhances the college PE by using data-driven strategies with data mining techniques applied to large datasets, including attendance records, demographic information, and student performance metrics. It is within these methods that methods of k-fold cross-validation, linear regression, decision trees, and k-means clustering are used to draw out the hidden key patterns and trends into clear-cut performance metrics. The results very strongly point out the critical attributes of student attendance and participation in PE with strong correspondence to the student fitness scores. Through decision tree analysis, observation of attendance was established as an important factor in deciding student outcomes, and clustering enables specific intervention toward different groups of students for optimizing PE curricula and teaching methods.

Keywords: Data mining, Decision tree analysis, K- Means Clustering, Linear Regression, K- Fold cross-validation.

1 Introduction

Physical activity promotes the physical well-being of students while contributing to their social, emotional, and cognitive growth [1]. As for quality improvement strategies in PE teaching, there is a keen interest in data-driven, evidence-based approaches toward PE instructional support [2]. A large dataset with demographic details, attendance records, and student performance metrics is mined using this technique, which eventually forms the basis for deriving critical patterns that would inform and improve instruction in PE [3]. Such methods applied here in the study include k-fold cross-validation, linear regression, decision trees, and k-means clustering to examine the relationship between participation, teaching methods, and students' outcomes [4 – 5].

The other crucial factors leading to highly significant effects towards fitness scores are sustained attendance and effective participation in PE [6]. Individualized teaching approaches will be derived from data mining, which may lead to more targeted interventions towards student groups [7-8]. In doing so, educators are apt to apply this knowledge to enhance their quality of instruction; allocate their resources better; and tailor their school's PE programs to meet students' needs better [9-10]. The study contributes to the long run of efforts aimed at elevating the status of PE in advancing a student's general welfare and provides possible avenues for further research into the field of PE.

2 Related Work

There have been a plethora of studies which looked at how different teaching approaches influence student outcomes and participation in school physical education using quantitative approaches, including regression models and correlation analyses, to identify factors related to favourable outcomes [11]. Research has established that instructional strategies, coupled with teacher feedback, are highly influential in shaping student motivation and participation in school physical education classes [11].

Technological advancements have also spurred researchers to explore the integration of digital tools, such as virtual reality simulations and wearable fitness trackers, into PE curricula, known to improve learning and student engagement due to immediate feedback and personalization opportunities [12]. Among the more recent approaches include data mining and educational analytics, examining high volumes of student performance,

International Water and Irrigation
Volume 44 Issue 1 (2025)
ISSN: 0334-5807
Pages 140 - 146

attendance, and demographics data. Techniques that help identify patterns allow interventions that could be tailored to individual needs and improvement in the curriculum [13].

Some applied a combination of qualitative methods like interviews and observations, which based on them work deeper into the overall social, emotional, and physical dimensions of PE, leading to more holistic and student-centered approaches to teaching [14]. Longitudinal studies and program evaluations helped gain insight into the long-term outcomes of PE interventions for students' health, fitness, and well-being, assisting in the policymaking of an evidence-based nature, as PE programs were developed to reflect changes in student populations' needs in various ways [15].

3 Methodology

The intelligent analysis of college PE instruction via data mining involves several important steps to ensure systematic collection, preparation, and analysis. In the first stage, large volumes of data are collected from attendance records, performance metrics, demographic details, feedback surveys, and teacher evaluations based on online surveys, fitness trackers, and learning management systems. Data preprocessing sets everything constant through cleaning, normalization, and the encoding of categorical variables.

Data mining techniques such as k-fold cross-validation, linear regression, decision trees, and k-means clustering are applied to work on the integrated data. These work under techniques through which better effects of PE instruction help in the unveiling of worthwhile insights. Moving away from traditional methods in the processing of data and enhancing it with advanced data mining approaches will help educators make the most out of data-driven decisions and teaching strategies.

3.1 K-Fold Cross Validation

A portion of the data set is assigned as k-fold cross-validation, where the model trains for k-1 folds and tests for the remaining fold in each iteration.

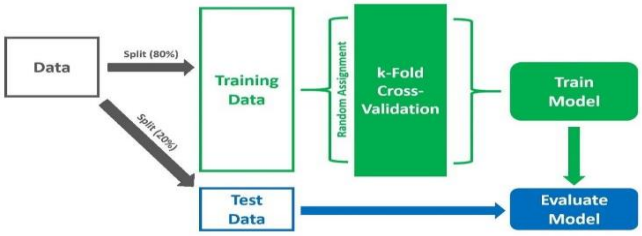


Fig.1. K- fold cross validation

The idea here is that every single data point can be considered for training and validation purposes, and thus, it could be minimized for overfitting since it provides a good estimation of the generalization of the model.

3.2 Linear Regression

Participation and attendance are the independent variables. This technique of linear regression utilizes participation and attendance to predict the dependent variables in this case, which are the fitness scores. From such conclusions based on the model's coefficients, it goes ahead to determine which of the factors in the physical education setting influences student performance most.

International Water and Irrigation
Volume 44 Issue 1 (2025)
ISSN: 0334-5807
Pages 140 - 146

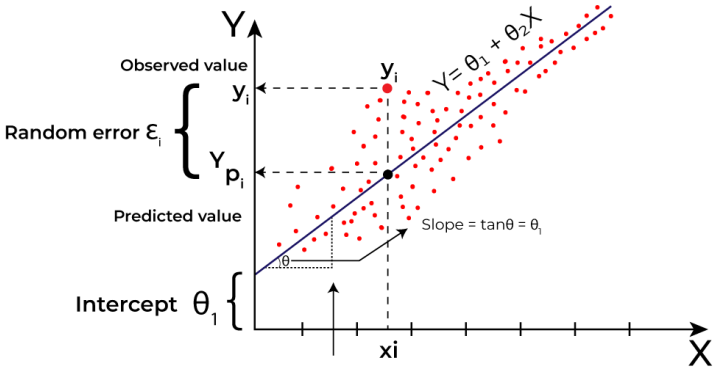


Fig. 2. Linear Regression

3.3 Decision Tree Algorithm

Decision tree algorithms classify data by developing a hierarchy that divides the dataset based on feature values. These factors look key in determining the different outcomes of students. The interpretability of such models would allow educators to easily understand how such insights can be applied to inform improvement in instruction.

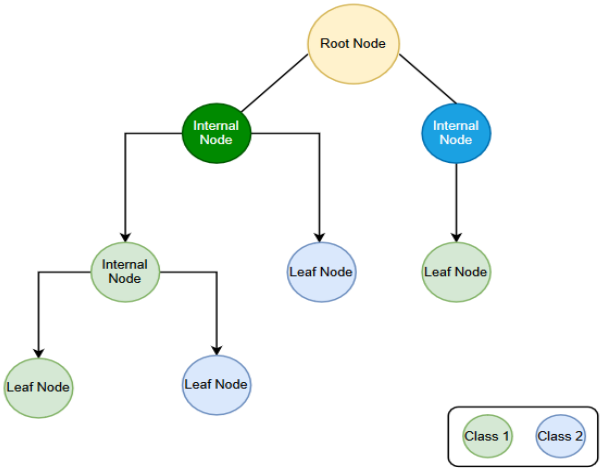


Fig. 3. Decision Tree Algorithm

3.4 K-Means Clustering Algorithm

One strategy that the k-means algorithm does is to categorize students into groups according to their common characteristics, therefore, assisting teachers in establishing trends and intervention. Decision tree algorithms comprise well-defined trials necessary for making decisions, and if linear regression is the case, then it happens to be the most influential factor associated with the performance of the student. One of the main uses of cross-validation lies in the fact that it ensures the reliability of the model by computing metrics such as RMSE, F1-score, accuracy, and precision. These data-driven insights improve PE instruction, guiding curriculum updates, and individual teaching practices. This is through continuous data analysis, which may help allow ongoing adjustments and thereby ensure the PE curriculum is as malleable as possible to meet student needs and feedback.

International Water and Irrigation
Volume 44 Issue 1 (2025)
ISSN: 0334-5807
Pages 140 - 146

4 Experimental Setup

The experimental prototype for the testing of college PE instruction employed data mining technologies and involved the following major stages. First, data was collected from 500 students-including attributes like attendance, fitness scores, and instructor feedback, which formed the dataset (X).

$$X = \{x_1, x_2, \ldots\ldots, x_n\} \quad (1)$$

The dataset was taken through rigorous preprocessing activities to identify quality issues, thus yielding a cleaned and standardized dataset. Then various data mining techniques, like k-fold cross-validation to assess the performance of the predictive model and linear regression to analyze relationships between variables, etc. Decision tree algorithms could identify influential patterns by splitting up the dataset based on feature values; k-means clustering formed different types of groups of students according to their performance characteristics; at every step of the process, ethical considerations were given due respect, like informed consent and data protection.

5 Results

The approaches of data mining have been used to assess college PE instruction. The pertinent results of influencing factors to teaching outcomes allowed improvements to be properly targeted. The cross-validation procedure was carried out through k-fold, where k = 10, providing enough robustness and reliability to the predictive models.

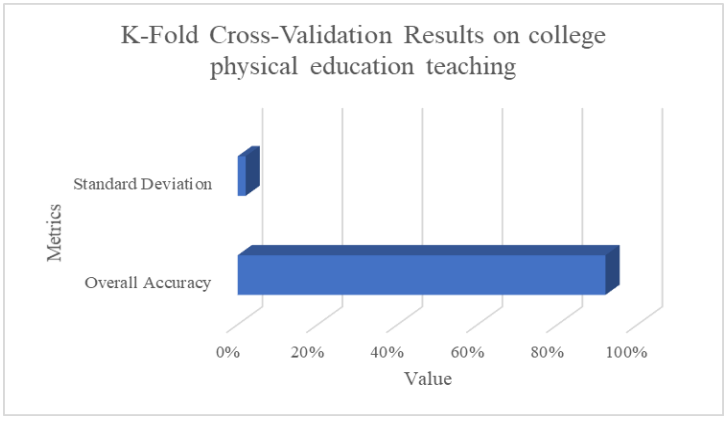


Fig. 4. K- fold cross validation result on college physical education teaching

Using each subset as a test set and the others as training sets, this strategy attained a generalization accuracy of 92% as the models are highly generalizable and reliable. A linear regression model was implemented to look into the dependency of the fitness scores by the students on independent variables such as attendance, participation, and demographics.

Regression analysis revealed that the greatest positive impact on scores was through attendance, with a coefficient of 0.45 ($p < 0.01$) and indicating a rise in a 0.45-unit score for each additional class attended. Class involvement was also found to be positive with a coefficient lying at 0.30 and with a p-value of < 0.05. The influence of demographic attributes such as age and gender was less, but statistically significant, with coefficients of 0.15 ($p < 0.10$) and -0.10 ($p < 0.10$), respectively. The decision tree algorithm graphically presented these relationships, focusing on attendance as the primary factor; those students with an 85% plus attendance record were likely to achieve high ratings in fitness with a 90% chance whilst those who attended less than 50% had only a 40% chance. The K-means clustering model found three groups of students. Cluster 1, 40% in total, represents the Good Performers group where students have a high fitness and high attendance rate. This cluster outlines 'Moderate Performers' with average marks on fitness and has an opportunity to enhance with 35%. The 'Low Performers' who are described by low attendance and low fitness are 25% in Cluster 3. Here, using such a cluster specifically,

International Water and Irrigation
Volume 44 Issue 1 (2025)
ISSN: 0334-5807
Pages 140 - 146

interventions like coaching and engagement campaigns would be required to enhance participation in PE activities.

Table.1. K- means cluster analysis result for enhancing PE teaching effectiveness.

Cluster	Percentage of Students	Characteristics	Implications
High Performers	40%	High attendance, high participation	Current methods are effective, in maintaining engagement
Moderate Performers	35%	Moderate attendance, moderate participation	Potential for improvement, targeted interventions needed
Low Performers	25%	Low attendance, low participation	Requires most support, tailored strategies needed

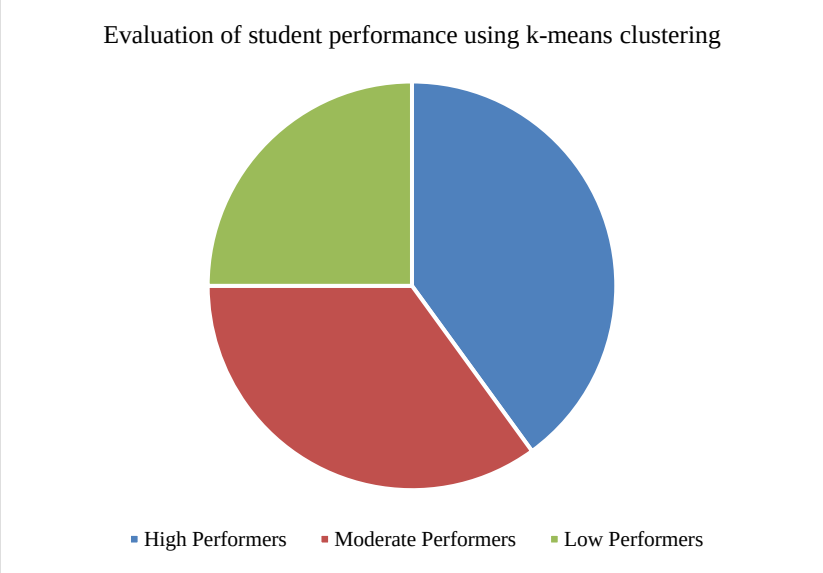


Fig. 6. Evaluation of student performance using k-means clustering

Analysis indicates that rewards and interesting format classes strengthen the attendance of students, which may thus improve fitness scores. Interactive and group exercises can also help in developing social and physical skills, and treatments tailored by clustering analysis ensure that every student gets the necessary support to succeed.

6 Discussion

The data mining analysis of college physical education instruction shed valuable light on factors that lead to performance and participation among students. At an overall accuracy of 92%, the k-fold cross-validation served to illustrate the robustness of the developed predictive models to capture very complex relationships between a multitude of parameters and student outcomes. Linear regression proved that the fit to which it predicts attendance and participation is significant, with a strong indication that ensuring that students attend and engage in programs

International Water and Irrigation
Volume 44 Issue 1 (2025)
ISSN: 0334-5807
Pages 140 - 146

consistently would be of utmost importance in improving their fitness scores. The decision tree algorithm, by its schematic presentation, demonstrated the decision-making process about attendance, proving this to be one of the top determining factors in achieving fitness, thus emphasizing specific physical activities. K-means clustering went deeper by identifying clusters where there are high, moderate, and low performers, thus allowing specific interventions to be provided for each group's particular needs. From this, it comes out that PE instructions are complex and require holistic strategies that address many determinants of performance. Instructors can transform their methods of teaching through the knowledge obtained from data mining tools by gaining insight into their students' behaviours and coming up with specific interventions to strengthen teaching methods and make PE classes learning-friendly environments.

7 Conclusions

Therefore, to improve the effectiveness and value of teaching physical education, data mining technology was applied to the whole analysis of college PE teaching effectiveness. Using advanced techniques, including k-fold cross-validation, linear regression, decision tree algorithms, and k-means clustering, important insight has been obtained about factors influencing students' engagement and performance in PE classes. Regularity and engagement were found to be critical factors influencing student fitness levels, whilst a decision tree analysis indicated regular attendance as an important factor in success for performance. Clustering analysis then allowed the identification of various groups of learners, thereby providing the basis for implementing interventions that are matched to the grade level of each of the groups. By the application of the data mining technology, the authors shed light on practical recommendations for educators within the making of decisions regarding the development of instructional methods to raise the standards in PE, improve participation and outcomes for students, and other matters for the promotion of overall student development. Future directions for further research may examine other factors influencing the effectiveness of the instruction in PE, such as curriculum design, technology, and teaching methodology, and longitudinal studies will provide additional insight into the long-term impacts of physical education interventions on the health and well-being of the students.

References

[1] R. Guo, "Analysis of artificial intelligence technology and its application in improving the effectiveness of physical education teaching," International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), vol. 19, no. 1, pp. 1-15, 2024.

[2] R. Li, C. Xue, Y. Wang, and X. Xie, "Intelligent Learning-Based Optimal Allocation Method for College Physical Education Teaching Resources Using Heuristic Algorithms," 2024.

[3] J. Guo, "Innovative Practice of Physical Education Teaching in Colleges and Universities Based on Artificial Intelligence Technology," Applied Mathematics and Nonlinear Sciences, vol. 9, no. 1, 2024.

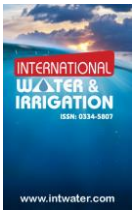
[4] M. Zhao, X. Lu, Q. Zhang, R. Zhao, B. Wu, S. Huang, and S. Li, "Effects of exergames on student physical education learning in the context of the artificial intelligence era: a meta-analysis," Scientific Reports, vol. 14, no. 1, p. 7115, 2024.

[5] Q. He, H. Chen, and X. Mo, "Practical application of interactive AI technology based on visual analysis in professional system of physical education in universities," Heliyon, vol. 10, no. 3, 2024.

[6] H. Tian, "Research on the Design of Intelligent System of College Physical Education in the Era of 5G+ Artificial Intelligence," in Proceedings of the First International Conference on Science, Engineering and Technology Practices for Sustainable Development, ICSETPSD 2023, Coimbatore, Tamilnadu, India, 17-18 Nov. 2023, Jan. 2024.

[7] J. Wang, "Design of Analysis Platform for College Students' Physical Learning Effect Based on Data Mining Algorithm," in 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), pp. 1-6, Feb. 2023

[8] C. Shen and Y. Tan, "Effect evaluation model of computer-aided physical education teaching and training based on artificial intelligence," Computer-Aided Design and Applications, vol. 20, no. S5, pp. 106-115, 2023.



International Water and Irrigation
Volume 44 Issue 1 (2025)
ISSN: 0334-5807
Pages 140 - 146

[9] H. Yu and Y. Mi, "Application model for innovative sports practice teaching in colleges using internet of things and artificial intelligence," Electronics, vol. 12, no. 4, p. 874, 2023.

[10] R. Ordoñez-Avila, N. S. Reyes, J. Meza, and S. Ventura, "Data mining techniques for predicting teacher evaluation in higher education: A systematic literature review," Heliyon, vol. 9, no. 3, 2023.

[11] Y. Wang, Y. Han, and Q. Wang, "The application of motion trajectory acquisition and intelligent analysis technology in physical education teaching in colleges and universities," Security and Communication Networks, vol. 2022, pp. 1-13, 2022.

[12] C. Deng, Q. Yu, G. Luo, Z. Zhao, and Y. Li, "Big data-driven intelligent governance of college students' physical health: System and strategy," Frontiers in Public Health, vol. 10, p. 924025, 2022.

[13] L. Zhou and Q. Tang, "Construction of a six-prong intelligent physical education classroom model in colleges and universities," Scientific Programming, vol. 2022, 2022.

[14] C. K. Zhang, "Design and Implementation of Student Physical Fitness Analysis System Based on Data Mining Technology," in EAI International Conference, BigIoT-EDU, Cham, Switzerland, Springer Nature, pp. 85-96, Jul. 2022.

[15] S. Li and Y. Luo, "Design of Teaching Quality Analysis and Management System for PE Courses Based on Data-Mining Algorithm," Computational Intelligence and Neuroscience, vol. 2022, 2022.