Research on Prediction of Online Training Effect of Newly Recruited PE Teachers in Junior and Senior Middle Schools Based on Machine Learning

Feng Zhao[1,3,a,*], Yuelin Yang[2,b], Luhan Xing[1,c]

[1]*School of Physical Education, Northeast Normal University, Changchun, China*
[2]*Department of Kinesiology, California Baptist University, California, US*
[3]*School of Physical Education and Health, Yili Normal University, Yili, China*
a. tiyuzhaofeng@163.com, b. 769006@calbaptist.edu, c. xingluhan032@163.com
corresponding author

Abstract: With the continuous development of online education, how to improve the teaching ability of new PE teachers through adaptive and effective online training has become an important research issue. Based on machine learning algorithm, this paper discusses the influence of different characteristics on the adaptability of online training for new physical education teachers, and evaluates the application of various models in predicting the training effect of teachers. The results show that factors such as teachers' professional experience, past training experience, school type and whether to participate in training with colleagues have a significant impact on the adaptability of online training. By comparing Logistic regression model, KNN model, random forest model, XGBoost model and support vector machine model, this paper finds that random forest model is the best in prediction accuracy and generalization ability. This study provides data support and theoretical basis for optimizing the online training of physical education teachers, and can provide reference for educational managers to formulate personalized training programs.

Keywords: New physical education teachers, Online training, Machine learning, Training adaptability.

1.   Introduction

With the rapid development of information technology, online education has become an indispensable education system in modern education. In this era of rapid network development, network training provides a flexible and efficient way for new teachers to learn, and it also provides an essential way for new teachers to improve their teaching ability and promote their professional development. However, although online training has many advantages, its adaptability and effectiveness are still affected by many factors, especially for PE teachers who need practice and interaction, whether online training can fully support the improvement of their teaching ability is still a question worth discussing.

New PE teachers are facing various challenges in the process of adapting to the educational environment, mastering teaching skills and improving classroom management ability. Under this background, how to evaluate the influence of different training methods on the teaching adaptability of physical education teachers has become an important topic to improve the quality of teachers'

education. Although the traditional face-to-face training can provide rich opportunities for interaction and practice, its coverage and flexibility are relatively low due to the limitation of time and space. Online training, on the other hand, provides more learning choices and space for teachers through flexible time arrangement and convenient learning platform,It also saves on additional costs such as transportation and lodging.

This study aims to explore the influence of different training methods, especially online training methods, on the teaching adaptability of new PE teachers. By collecting data including gender, age, educational background, training duration and other characteristics, combined with a machine learning algorithm for data modeling and analysis, this study tries to reveal the key factors affecting the training effect of teachers, and provide theoretical support and practical guidance for online education. Through this study, we hope to provide data support for the online training mode of physical education teachers, and provide a reference for the innovation of teacher training mode in other disciplines.

2.   Literature Review

With the rise of online education and advancements in educational technology and big data, machine learning is increasingly applied to enhance teaching quality and personalized learning. In physical education, improving new teachers' adaptability through online training has become a pressing challenge. This study uses machine learning to analyze the impact of different training methods on the adaptability of junior and senior high school PE teachers, examining the predictive role of related variables on training outcomes.

Online education's flexibility and convenience make it essential to modern learning. Zhou Hongwei highlighted that adaptive online learning models, integrating big data, can personalize learning paths and improve outcomes [1]. Educational big data precisely analyzes behaviors and adjusts strategies to enhance learning, aligning with this study's focus. Using machine learning, this research aims to optimize online PE teacher training by evaluating adaptability and refining training methods through data-driven insights.

Villegas-Ch et al [2]. further discussed the integrated application of machine learning and data analysis in learning management system (LMS), and pointed out that through machine learning technology, teachers can evaluate students' learning progress more accurately and adjust teaching strategies in real time. By integrating machine learning algorithm in LMS, the interactivity and intelligence of online education have been improved, which is very important for personalized guidance in teacher training. This paper also uses machine learning technology to predict the training adaptability of new physical education teachers in order to provide personalized teaching suggestions and support.

Machine learning is widely used in the field of education, especially in the prediction of students' grades and the evaluation of teaching effect. Zhai et al [3]. reviewed the application of machine learning in science education evaluation, and pointed out that machine learning method can effectively process large-scale education data and find potential learning patterns and evaluation errors. In addition, the meta-analysis of Zhai et al [4]. shows that scientific evaluation based on machine learning can improve the consistency of human and machine scores and provide more accurate tools for educational evaluation. This research provides strong support for this paper, especially in predicting the effect of teacher training by using machine learning.

Luan and Tsai highlighted machine learning's role in precision education, emphasizing personalized content creation by analyzing student characteristics [5]. Machine learning explores learning habits and behaviors, tailoring teaching plans to improve outcomes, aligning with this study's goal of predicting new PE teachers' adaptability in online training based on factors like gender, age, and education.

Alhothali et al [6]. reviewed machine learning applications for predicting student achievements in online courses, noting the varying strengths of algorithms like Decision Trees, SVM, and Random Forest. By analyzing participation, learning time, and submissions, these methods predict outcomes to guide teaching. Similarly, this study employs machine learning models, including SVM, Random Forest, and XGBoost, to predict PE teachers' training effectiveness, providing actionable insights for improved training strategies.

The integration of online education and machine learning has revolutionized education by leveraging big data for teacher training, learning outcome prediction, and strategy adjustment. This study applies machine learning to analyze how teacher characteristics impact training adaptability, offering theoretical and practical guidance for optimizing online PE teacher training. Building on prior research, it provides data-driven insights for enhancing personalization and accuracy. Future research can explore additional variables and refine models to further improve online training effectiveness.

3. Data and Model

3.1. Data introduction

In machine learning, features are variables that describe the attributes of data samples and serve as the fundamental input for prediction or classification models. These features act as independent variables within the model to capture differences and patterns among samples. As such, feature selection and extraction are crucial to the model's performance. High-quality features can significantly enhance both the accuracy and generalization ability of the model [7].

This study focuses on the adaptability and training effectiveness of new PE teachers in junior and senior high schools amidst evolving educational environments. Data was collected through questionnaires, interviews, and school records, covering characteristics such as gender, age, educational background, professional experience, past training participation, school type, and participation with colleagues. These features reflect teachers' backgrounds, work environments, and training adaptability.

Key variables include gender and age, providing demographic insights into differences in training adaptability, and educational background, highlighting the impact of qualifications and majors on training outcomes. Particular emphasis is placed on how a physical education background influences the application of teaching methods in online and face-to-face training. By analyzing these factors, the study aims to understand and improve new PE teachers' adaptability to various training methods, offering guidance for future educational training strategies.

Professional experience: Teachers' teaching experience has a far-reaching impact on their career development. This study explores whether professional experience can help teachers better adapt to different training methods by recording the teaching years and practical experience of new teachers.

Past training participation records: collect whether teachers have participated in relevant vocational training to evaluate their acceptance and adaptability to new training forms.

Types of schools: including public schools and private schools. The teaching resources, students' backgrounds and cultural atmosphere of different types of schools may affect teachers' training needs and adaptability. By analyzing the training effect of teachers in different types of schools, it can provide a basis for the design of training courses.

Colleagues participate in training together: It is found that peer support among teachers has a positive impact on the training effect. Therefore, investigating whether teachers participate in training with colleagues has become an important variable in this study.

Training duration: As an important factor affecting the training effect, training duration reflects whether teachers have enough time to learn and adapt to new teaching methods, especially in online

training mode, the allocation of learning time and teachers' time management ability have become important factors affecting adaptability.

3.2. Model introduction

Logistic Regression is a binary classification method that estimates probabilities using a logistic function. Simple and interpretable, it aids decision-making by identifying influential factors. In this study, it evaluates how various features impact new PE teachers' training adaptability, helping educators predict and enhance training outcomes effectively.

The K-Nearest Neighbors (KNN) algorithm classifies data based on the majority vote of its k nearest neighbors, offering simplicity and ease of implementation [8]. In this study, KNN predicts new PE teachers' adaptability to online training by analyzing similarities with successfully adapted teachers, providing insights into training outcomes.

Random Forest is a robust ensemble method for binary classification, constructing multiple decision trees and aggregating their outputs to reduce overfitting [9]. It effectively handles diverse features like gender, age, and educational background, while ranking feature importance to identify key factors influencing online training adaptability.

XGBoost, renowned for its efficiency and accuracy, excels in handling complex, high-dimensional data [10]. Its advanced gradient boosting implementation, regularization features, and ability to manage missing data ensure precise predictions and strong generalization. Both models offer valuable tools for analyzing factors affecting the effectiveness of training methods for new PE teachers.

Support Vector Machine (SVM) excels in binary classification by maximizing the margin between classes, effectively handling high-dimensional data [11]. In this study, SVM identifies boundaries between successful and unsuccessful adaptations to online training, leveraging kernel functions to capture complex patterns in teacher adaptation data for precise predictions.

Each of these models brings unique strengths to the binary classification task at hand, offering different perspectives and insights into the factors influencing the effectiveness of online training methods for new physical education teachers. By leveraging these models, we can obtain a comprehensive understanding of the key determinants of training success and tailor strategies to improve adaptation outcomes.

4. Model results

The results from the models reveal varying levels of performance in predicting the effectiveness of training methods for new physical education teachers. Logistic Regression, a simpler model, achieved an accuracy of 0.548, which suggests it may struggle with capturing complex patterns in the data. Its precision and recall scores indicate moderate effectiveness, but it might not be the best choice for this task.

The K-Nearest Neighbors (KNN) model showed improvement with an accuracy of 0.663. Both precision and recall scores are slightly above the accuracy, indicating that this model is better at identifying true positives compared to Logistic Regression. However, it still leaves room for improvement in overall performance.

The Random Forest model demonstrated impressive results with an accuracy of 0.826, making it the most accurate among the tested models. Its precision and recall scores are also high, suggesting that it balances sensitivity and specificity well. This model's performance indicates it can effectively handle the complexity and variability in the data.

Table 1: Classifiers Accuracy Score

Model	Accuracy	Precision	Recall
LogisticRgeress	0.548	0.602	0.654
KNN	0.663	0.682	0.680
RF	0.826	0.782	0.805
XGBoost	0.791	0.734	0.752
SVM	0.802	0.761	0.789

XGBoost also performed well with an accuracy of 0.791. While slightly lower than Random Forest, its precision and recall scores reflect a robust ability to predict outcomes accurately. XGBoost's advanced techniques contribute to its strong performance, making it a reliable choice for this task.

Support Vector Machine (SVM) achieved an accuracy of 0.802, which is competitive with XGBoost and slightly below Random Forest. Its precision and recall scores suggest that it is effective in classifying the data, offering a good balance between identifying true positives and minimizing false positives. This performance highlights SVM's capability in managing the complexities of the training data.

Overall, Random Forest appears to be the most suitable model for this task, given its highest accuracy and balanced precision and recall scores, closely followed by SVM and XGBoost. These models demonstrate strong potential for accurately predicting the effectiveness of online training methods for new physical education teachers.
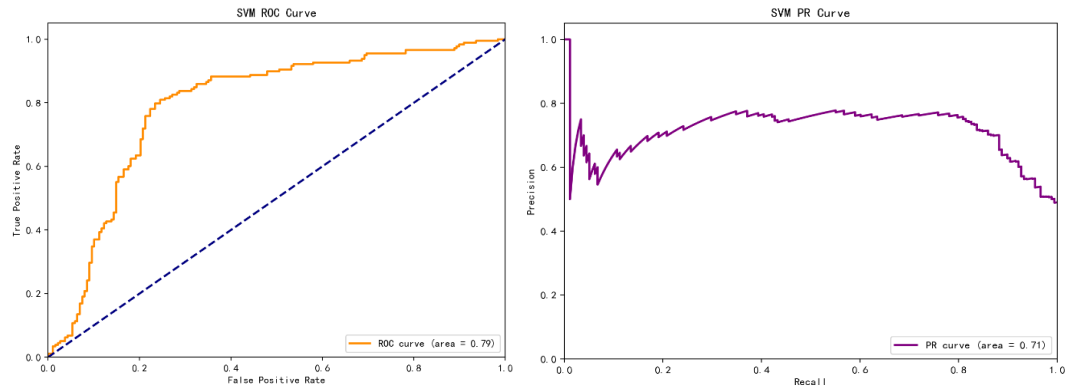

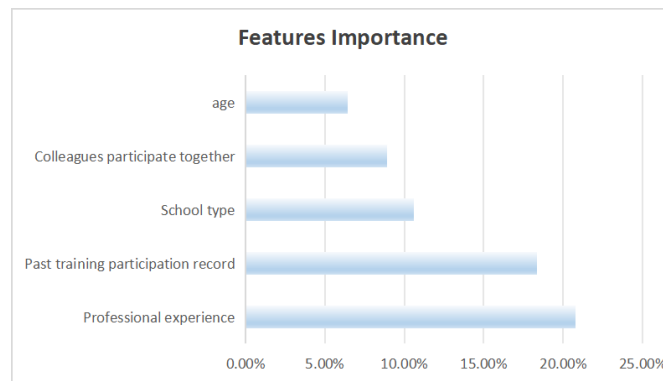Figure 1: ROC diagram and PR diagram of RandomForest Model

Figure 2: Feature Importance Ranking

Professional experience (20.80%) is the most significant factor in predicting PE teachers' adaptability to online training. Experienced teachers adapt better to practical teaching challenges, while less experienced teachers may benefit from tailored training with basic modules and case studies. Past training participation (18.37%) also strongly impacts adaptability, as prior training helps teachers acclimate to new methods. Introductory guidance for untrained teachers can improve their familiarity with online formats.

School type (10.61%) influences adaptability due to differences in resources and environments. Training programs should consider these disparities and adjust content to suit diverse school settings. Peer participation (8.92%) enhances adaptability through group learning and support. Encouraging study groups and interactive discussions can foster collaboration and experience-sharing among teachers.

While age (6.45%) has a smaller impact, older teachers may face challenges with technology. Simplified technical training, video tutorials, and personalized support can help them adapt more effectively. These findings highlight the need for adaptive, tailored online training programs to address the varying needs of PE teachers, improving overall training effectiveness.

5. Conclusions

This study examined the effectiveness of training methods, particularly online training, for new physical education (PE) teachers. Using machine learning algorithms, key factors like gender, age, educational background, professional experience, past training participation, and school type were analyzed to predict teachers' adaptability to training methods. The results offer valuable insights into improving online training for PE teachers.

The Random Forest model was the most effective, achieving an accuracy of 0.826, with balanced precision and recall, making it the most reliable tool for identifying adaptability factors. Support Vector Machine (SVM) and XGBoost also performed well, with accuracies of 0.802 and 0.791, respectively. Logistic Regression and K-Nearest Neighbors (KNN) showed lower performance, with Logistic Regression struggling to handle data complexity.

Feature importance analysis revealed professional experience as the top predictor of adaptability, with experienced teachers adapting better to online formats. Tailored training content for less experienced teachers, including foundational knowledge and practical cases, is recommended. Past training participation also significantly impacted adaptability, with prior training increasing the likelihood of success. For teachers without training experience, introductory resources and personalized guidance are suggested.

School type influenced adaptability due to varying resources and environments, emphasizing the need for tailored training programs. Peer participation enhanced learning through collaboration and support, suggesting the inclusion of group discussions or study groups. While age was less significant, older teachers faced challenges with technology. User-friendly interfaces and support like tutorials can help. These findings provide a foundation for designing adaptive online training methods to support diverse educator needs.

References

[1] Zhou Hongwei. Research on the Adaptive Learning Model of Online Education Based on Educational Big Data and Its Application [J]. Research on Continuing Education, 2023, (03): 110-114.
[2] Villegas-Ch W, Román-Cañizares M, Palacios-Pacheco X. Improvement of an online education model with the integration of machine learning and data analysis in an LMS[J]. Applied Sciences, 2020, 10(15): 5371.
[3] Zhai X, Yin Y, Pellegrino J W, et al. Applying machine learning in science assessment: a systematic review[J]. Studies in Science Education, 2020, 56(1): 111-151.
[4] Zhai X, Shi L, Nehm R H. A meta-analysis of machine learning-based science assessments: Factors impacting machine-human score agreements[J]. Journal of Science Education and Technology, 2021, 30: 361-379.

[5] Luan H, Tsai C C. A review of using machine learning approaches for precision education[J]. Educational Technology & Society, 2021, 24(1): 250-266.

[6] Alhothali A, Albsisi M, Assalahi H, et al. Predicting student outcomes in online courses using machine learning techniques: A review[J]. Sustainability, 2022, 14(10): 6199.

[7] Hosmer Jr D W, Lemeshow S, Sturdivant R X. Applied logistic regression[M]. John Wiley & Sons, 2013.

[8] Peterson L E. K-nearest neighbor[J]. Scholarpedia, 2009, 4(2): 1883.

[9] Biau G, Scornet E. A random forest guided tour[J]. Test, 2016, 25: 197-227.

[10] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.

[11] Jakkula V. Tutorial on support vector machine (svm)[J]. School of EECS, Washington State University, 2006, 37(2.5): 3.