Research on Efficiency Improvement and Empirical Study of Educational Resource Allocation Based on Improved Machine Learning Algorithm

Ruijun Ban
School of Marxism, Shaoxing University, Shaoxing 312000, Zhejiang, China
Corresponding author: Ruijun Ban (e-mail: 15532665836@163.com)

ABSTRACT Balanced and precise educational resource allocation is a core issue in advancing educational equity and improving overall quality. To address this problem, this paper proposes an optimized educational resource allocation model based on an improved random forest algorithm. First, theories related to educational resource allocation and machine learning are reviewed, and a core indicator system affecting resource allocation efficiency is clarified. Second, to address bias in feature weight allocation in traditional random forest algorithms, an adaptive weighting mechanism and cross-validation optimization strategy are introduced to construct an improved random forest algorithm. Third, basic educational resource data are used as samples, and the improved algorithm is applied to predict and optimize resource allocation efficiency. Comparative experiments are conducted to verify the model's superiority. Finally, targeted suggestions for improving allocation efficiency are proposed based on empirical results. The results show that the R^2 value of the improved random forest algorithm is 5.3 percentage points higher than that of the traditional algorithm, increasing from 0.873 to 0.926, with a 12.3% reduction in RMSE. The method effectively identifies key bottlenecks and supports precise resource optimization.

INDEX TERMS machine learning, educational resource allocation, random forest, efficiency prediction, resource optimization

I. INTRODUCTION

A. RESEARCH BACKGROUND

THE development of basic education and the thorough implementation of quality-oriented education are prominent issues confronting China's education sector currently [1-3]. Unbalanced allocation and low efficiency in education persist, with the dual gaps tending to widen, which has become a prominent contradiction hindering the connotative development of education [4-6].

Traditional educational resource allocation models are mostly based on administrative orders or empirical judgments, lacking accurate analysis of the matching degree between resource supply and demand, resulting in low resource allocation efficiency [7]. Machine learning algorithms, relying on their powerful data mining and predictive analysis capabilities, have provided new ideas for the optimization of educational resource allocation [8,9]. Through in-depth analysis of educational resource-related data using machine learning algorithms, it is possible to accurately identify resource demand hotspots, predict resource allocation gaps, thereby realizing precise resource investment and improving resource allocation efficiency [10,11]. Therefore, this paper intends to utilize improved machine learning algorithms to improve efficiency, mirroring the precise automated sorting and grading processes essential to high-quality manufacturing.

B. RESEARCH STATUS

Presently, there are relatively few studies integrating educational resource allocation and machine learning. In the field of educational resource allocation, research primarily focuses on the issue of resource balance between regions and urban-rural areas, employing various methods to evaluate resource allocation efficiency [12,13]. For instance, certain methods are utilized to measure the efficiency of basic educational resource allocation, indicating that the regional structure of educational investment is a key factor influencing efficiency [14-16]. In terms of the integration research of machine learning and educational resource allocation, domestic scholars have begun to try to apply machine learning algorithms to resource allocation optimization. For example, the traditional random forest algorithm is used to predict educational resource demand [17,18], but this research does not improve the algorithm, the prediction accuracy needs to be improved, and there is a lack of in-depth analysis of empirical results and policy recommendations. In summary, although existing studies have paid attention to the application of machine learning algorithms in educational resource allocation, there are still the following deficiencies:

Firstly, traditional machine learning algorithms have biases in feature weight allocation, leading to insufficient prediction accuracy; Secondly, there is a lack of in-depth analysis of the mechanism through which algorithms improve resource allocation efficiency; Thirdly, the sample coverage and data timeliness of empirical research need to be improved.

Based on this, this paper constructs an optimized model for educational resource allocation based on advanced machine learning algorithms, conducts large-sample empirical research, and makes up for the deficiencies of existing research.

C. RESEARCH SIGNIFICANCE

1) Theoretically, this approach enriches the application of educational resource allocation theory and machine learning within the engineering sector, providing a robust interdisciplinary framework for optimizing specialized technical training and talent cultivation;

2) Practically, it offers technical support for educational management departments to formulate scientific and reasonable policies, and promote the transformation of educational resource allocation towards refinement and high efficiency.

D. RESEARCH CONTENT AND METHODS

1) Research Content

This research mainly includes the following aspects:

1) Systematically sort out the relevant theories of educational resource allocation and machine learning, laying the theoretical foundation for the research;

2) Construct a comprehensive evaluation index system for educational resource allocation efficiency, and screen the core feature variables affecting resource allocation efficiency;

3) Improve the random forest algorithm to construct an advanced random forest algorithm suitable for educational resource allocation efficiency prediction;

4) Select basic educational resource data as samples to conduct empirical research and verify the effectiveness of the model;

5) Propose targeted policy recommendations based on the empirical results.

2) Research Methods

(1) Literature review method:

Systematically sort out the literature in related fields such as educational resource allocation and machine learning algorithms, clarify the research status and development trends, and lay a theoretical foundation for the research;

(2) Index system construction method:

Combined with the actual needs of educational resource allocation, construct an evaluation index system from three dimensions: resource supply, resource demand, and resource allocation environment;

(3) Algorithm improvement method:

Aiming at the deficiencies of the traditional random forest algorithm, introduce an adaptive weighting mechanism and cross-validation optimization strategy to improve the performance of the algorithm;

(4) Empirical research method:

Take the basic educational resource data from 2018 to 2022 as research samples, use the advanced algorithm to conduct empirical analysis, and verify the effectiveness of the model.

E. RESEARCH INNOVATIONS

(1) Algorithm innovation:

Aiming at the bias problem in feature weight allocation of the traditional random forest algorithm, an adaptive weighting mechanism is introduced to dynamically adjust the weights according to the importance of features;

(2) Research perspective innovation:

From the perspective of matching resource supply and demand, construct a comprehensive evaluation index system for educational resource allocation efficiency, breaking through the limitation of traditional research that only focuses on resource supply;

(3) Empirical innovation:

Select the latest panel data of 31 provinces in China from 2018 to 2022 for empirical research. The sample coverage is wide and the data timeliness is strong, so the research conclusions are more convincing and practical.

II. RELATED THEORETICAL BASIS

A. EDUCATIONAL RESOURCE ALLOCATION THEORY

Educational resource allocation refers to allocating educational resources obtained through various channels to various components of education in a certain allocation method to form a certain resource structure. The optimization of resource structure is the destination of educational resource allocation issues. Essentially, educational resource allocation involves the distribution of resources among various components or different subsystems within the education system, which not only includes the allocation of total social resources to the education field but also the distribution of educational resources among different regions and schools. Educational resource allocation theory mainly includes equity theory, efficiency theory and demand-oriented theory:

- Equity theory emphasizes that educational resources should be allocated equally ensure that every educated person enjoys equal educational opportunities;
- Efficiency theory focuses on the allocation efficiency of educational resources, pursuing the maximum educational output with the minimum resource input;
- Demand-oriented theory holds that educational resource allocation should be centered on the actual

needs of educated people to achieve precise resource supply.

The efficiency evaluation of educational resource allocation is an important standard to measure the rationality of resource allocation, mainly including three dimensions: allocation efficiency, use efficiency and output efficiency:

- Allocation efficiency focuses on the rationality of resource allocation among different fields and regions;
- Use efficiency focuses on the waste degree of resources in the use process;
- Output efficiency focuses on the proportional relationship between resource input and educational output.

This paper focuses on allocation efficiency, optimizes resource allocation schemes through improved machine learning algorithms, and improves the rationality and accuracy of resource allocation.

B. MACHINE LEARNING RELATED THEORIES

Both classification and regression methods in machine learning aim to obtain a real-valued function through training samples. The essence of regression is to calculate the continuous output y value through the training data set, that is, for any input x, solve the corresponding y value using the decision function. The essence of classification is to classify each sample into its corresponding discrete category through the training set, that is, for any input x, solve the corresponding category through the decision function. The core difference lies in the value range of the output: classification outputs discrete category labels, while regression outputs continuous real numbers. Both technologies have been widely applied in prediction and optimization problems.

As we all know, random forest (RF) [19-21] aggregates information during prediction, so it can obtain good prediction results. In the decision-making process, random forest can also generate a large number of variables, thereby avoiding overfitting of the model. However, the traditional random forest algorithm adopts an equal weight method in feature weight allocation, failing to consider the differences in the importance of different features, which leads to insufficient prediction accuracy and stability of the algorithm when processing high-dimensional data.

C. INTEGRATION MECHANISM OF MACHINE LEARNING AND EDUCATIONAL RESOURCE ALLOCATION

1) Data collection and preprocessing: collect data related to educational resource supply, demand, allocation environment, etc., and perform preprocessing such as cleaning and standardization;

2) Feature extraction and selection: extract feature variables affecting resource allocation efficiency from the preprocessed data, and screen out core features;

3) Model construction and optimization: use machine learning algorithms to construct a prediction model for educational resource allocation efficiency, and improve model performance by optimizing algorithm parameters.

III. CONSTRUCTION OF EDUCATIONAL RESOURCE ALLOCATION EFFICIENCY EVALUATION INDEX SYSTEM

A. CONSTRUCTION PRINCIPLES

To construct a scientific and reasonable evaluation index system for educational resource allocation efficiency, this paper follows the following construction principles:

1) Scientific principle: the selection of indicators should be based on the relevant theories of educational resource allocation, conform to the objective reality of educational resource allocation in China, and ensure the scientificity and rationality of the index system;

2) Comprehensive principle: the index system should cover multiple dimensions such as resource supply, resource demand, and resource allocation environment, and comprehensively reflect the factors affecting resource allocation efficiency;

3) Operability principle: the data of indicators should be easy to obtain, and can be collected through channels such as statistical yearbooks and public data of educational departments;

4) Targeted principle: The selection of indicators should focus on the actual situation of basic educational resource allocation in China, highlighting the key issues and core factors in the allocation process.

B. CONSTRUCTION OF INDEX SYSTEM AND VARIABLE EXPLANATION

Combined with the above construction principles, this article constructs an evaluation index system from three dimensions: resource supply, resource demand, and resource allocation environment. Prior to model training, a collinearity diagnosis was conducted using variance inflation factor (VIF) analysis. Variables with VIF > 10 were considered highly collinear and excluded. As a result, "Regional GDP (X11)" and "Per capita disposable income (X12)" were removed from the final feature set due to their high collinearity with "Public financial investment (X4)" (VIF values of 12.8 and 11.5, respectively). and selects a total of 13 feature variables, as follows:

1) Resource supply dimension: reflects the supply capacity of educational resources.

- X1: Number of full-time teachers
- X2: Per-student teaching instruments and equipment value
- X3: Per-student building area of school buildings
- X4: Public financial investment in education
- X5: Book collection

2) Resource demand dimension: reflects the demand scale and demand structure of educational resources.

- X6: Number of students in school
- X7: School enrollment rate of school-age children
- X8: Proportion of left-behind children
- X9: Urban-rural student ratio
- X10: Proportion of ethnic minority students

3) Resource allocation environment dimension: reflects the external environment affecting educational resource allocation.

- X11: Urbanization rate
- X12: Traffic accessibility
- X13: Quality of educational administrative personnel

This paper selects educational resource allocation efficiency as the explained variable (Y), and uses the DEA method to evaluate the educational resource allocation efficiency value of each province as the output variable of the model.

The specific measurement process is as follows: take the 5 variables in the resource supply dimension as input indicators, and take educational output indicators such as per-student average score and graduation rate as output indicators, and use the DEA-CCR to measure the efficiency value of each province. The efficiency value is between 0 and 1, and the closer it is to 1, the higher the resource allocation efficiency.

C. DATA SOURCES AND PREPROCESSING

The panel data of 31 provinces in China from 2018 to 2022 was selected as research samples. The data are obtained from the "China Education Statistics Yearbook", "China Statistical Yearbook" and the statistical bulletins on the development of education undertakings of various provinces.

The collected data are preprocessed as follows:

(1) Missing value processing: for a small amount of missing data, linear interpolation is used to supplement; for data with more missing values, the corresponding samples are eliminated;

(2) Outlier processing: the Z-score standardization method was used to identify outliers. First, the Z-score of each data point was calculated as Z = (X - μ)/σ, where μ is the variable mean and σ is the standard deviation. Outliers with |Z| > 3 were replaced with the variable's median (more robust than the mean for skewed data) to minimize data distortion;

(3) Data standardization: to eliminate the dimensional differences between different variables, the min-max standardization method is used to standardize all variables to the [0, 1] interval. The preprocessed data set contains a total of 155 samples (31 provinces $\times$ 5 years).

IV. CONSTRUCTION OF OPTIMIZED MODEL FOR EDUCATIONAL RESOURCE ALLOCATION EFFICIENCY BASED ON IMPROVED RANDOM FOREST ALGORITHM

A. PRINCIPLE OF TRADITIONAL RANDOM FOREST ALGORITHM

(1) Bootstrap sampling: randomly select n samples from the original data set to construct multiple training sets;

(2) Decision tree construction: for each training set, randomly select m feature variables, and construct a decision tree based on information gain or Gini coefficient;

(3) Ensemble prediction: the prediction results of multiple decision trees are aggregated through voting or weighted averaging to obtain the final prediction outcome.

The traditional random forest algorithm has high prediction accuracy, but it adopts an equal weight method in feature weight allocation, without considering the differences in the importance of different features, which leads to the problem of insufficient prediction accuracy when the algorithm processes high- dimensional data.

B. CONSTRUCTION OF ADVANCED RANDOM FOREST ALGORITHM

Aiming at the deficiencies of the traditional random forest algorithm, this paper introduces an adaptive weighting mechanism and cross-validation optimization strategy to construct an advanced random forest algorithm.

(1) Feature importance evaluation:

The Gini coefficient method is used to calculate the importance score of each feature variable. The calculation formula of the Gini coefficient is:

$$\text{Gini} = 1 - \sum (p_i)^2 \tag{1}$$

Where, p_i is the proportion of the i-th class of samples in the node. The larger the Gini coefficient value, the higher the importance of the feature.

(2) Introduction of adaptive weighting mechanism:

The adaptive weighting mechanism adjusts the feature sampling probability during decision tree construction based on Gini importance scores. Specifically, the weight of each feature w_i is calculated as the normalized Gini importance score (Formula 2), and this weight is incorporated into the feature selection probability distribution for each tree split. This ensures that features with higher importance are more likely to be selected during tree construction, rather than applying post-hoc weighted averaging of prediction results. The calculation formula of the feature weight is:

$$w_i = \frac{S_i}{\sum S_i} \tag{2}$$

Where, w_i is the weight of the i-th feature, and S_i is the importance score of the i-th feature. Through the adaptive weighting mechanism, features with higher importance play a greater role in the decision tree construction process.

(3) Cross-validation parameter optimization:

The 5-fold cross-validation method is used to optimize the algorithm parameters. The preprocessed dataset is randomly divided into 5 subsets, with each subset serving as a validation set in turn, and the remaining 4 subsets as a training set. Training and validation are repeated 5 times, and the parameter combination with the smallest prediction error is selected as the optimal parameter.

(4) Ensemble prediction:

Construct multiple weighted decision trees based on the optimized parameters, and perform a weighted average of the prediction results of each weighted decision tree to obtain the final predicted value of educational resource allocation efficiency.

C. MODEL PERFORMANCE EVALUATION INDICATORS

This paper selects MAE, MSE, RMSE and R^2 as performance evaluation indicators.

The calculation formulas of each indicator are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum |y_i - \hat{y}_i| \tag{3}$$

$$\mathrm{MSE} = \frac{1}{n}\sum (y_i - \hat{y}_i)^2 \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum (y_i - \hat{y}_i)^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \tag{6}$$

V. EMPIRICAL RESEARCH

A. EXPERIMENTAL DESIGN

This paper designs a comparative experiment, comparing the advanced random forest algorithm with the traditional random forest algorithm, SVM algorithm, and BP algorithm.

(1) Data set division: Divide the preprocessed 155 samples into a training set (108 samples) and a test set (47 samples) according to the ratio of 7:3;

(2) Parameter setting:

- For the improved random forest algorithm, the optimal number of decision trees is determined to be 100 and the number of feature selections is 8 through 5-fold cross-validation;
- The parameters of the traditional random forest algorithm are consistent with the improved algorithm, except that the adaptive weighting mechanism is not introduced;
- The SVM algorithm uses a radial basis kernel function with a penalty coefficient C=1.0;
- The BP neural network algorithm has 10 hidden layer nodes, a learning rate of 0.01, and an iteration number of 1000;

(3) Model training and prediction:

Train and predict using the four algorithms respectively, and calculate the performance evaluation indicators of each algorithm;

(4) Stability verification:

Repeat the experiment 10 times, calculate the mean and standard deviation of the performance evaluation indicators of each algorithm, and verify the stability of the algorithm.

B. EXPERIMENTAL RESULTS AND ANALYSIS

1) Comparative Analysis of Different Algorithms' Performance

The results of the performance evaluation indicators of the four algorithms are shown in Table 1. It can be seen from Table 1 that the MAE, MSE, and RMSE of the improved random forest algorithm are 0.032, 0.0015, and 0.038 respectively; the R^2 of the improved random forest algorithm is 0.926, which is higher than the other three algorithms.

TABLE 1. Comparison of Performance Evaluation Indicators of Different Algorithms

Algorithm	MAE	MSE	RMSE	R^2
Improved Random Forest	0.032	0.0015	0.038	0.926
Traditional Random Forest	0.045	0.0028	0.053	0.873
SVM	0.058	0.0042	0.065	0.815
BP Neural Network	0.063	0.0048	0.069	0.792

This makes clear that the fitting effect of the improved random forest algorithm are better than those of the traditional algorithms, and the introduction of the adaptive weighting mechanism and cross-validation optimization strategy can effectively improve the performance of the algorithm.

2) Analysis of Algorithm Stability Verification

After repeating the experiment 10 times, the standard deviations of the performance evaluation indicators of the four algorithms are shown in Table 2. It can be seen from Table 2 that the standard deviations of MAE, MSE, and RMSE of the improved random forest algorithm are 0.002, 0.0001, and 0.003 respectively, which are all lower than those of the other three algorithms; the standard deviation of R^2 of the improved random forest algorithm is 0.012, which is lower than the other three algorithms.

TABLE 2. Stability Verification Results of Different Algorithms

Algorithm	MAE	MSE	RMSE	R^2
Improved Random Forest	0.032 ± 0.002	0.0015 ± 0.0001	0.038 ± 0.003	0.926 ± 0.012
Traditional Random Forest	0.045 ± 0.004	0.0028 ± 0.0003	0.053 ± 0.005	0.873 ± 0.021
SVM	0.058 ± 0.006	0.0042 ± 0.0005	0.065 ± 0.007	0.815 ± 0.028
BP Neural Network	0.063 ± 0.007	0.0048 ± 0.0006	0.069 ± 0.008	0.792 ± 0.032

This indicates that the stability of the improved random forest algorithm is better than that of the traditional algorithms, and the introduction of the cross-validation optimization strategy can effectively improve the stability of the algorithm.

3) Cross-Validation Variance Analysis

To address the issue of unreported cross-validation variance, Table 3 presents the full 5-fold cross-validation results for the improved random forest algorithm, including mean values and standard deviations across folds. The low variance (e.g., R^2 standard deviation = 0.018 across folds) indicates that the model's performance is consistent and reliable despite the small dataset size.

TABLE 3. 5-Fold Cross-Validation Results of Improved Random Forest Algorithm

Fold	MAE	MSE	RMSE	R^2
1	0.031	0.0014	0.037	0.932
2	0.033	0.0016	0.040	0.921
3	0.030	0.0013	0.036	0.940
4	0.034	0.0017	0.041	0.915
5	0.032	0.0015	0.038	0.922
Mean±SD	0.032±0.0015	0.0015±0.00016	0.038±0.0021	0.926±0.018

4) Feature Importance Analysis

From Table 3, we can see that the top 5 feature variables in terms of importance scores are X4, X1, X6, X11, and X2, which are the core feature variables affecting educational resource allocation efficiency.

- Public financial investment in education has the highest importance score, indicating that capital investment is a key factor affecting resource allocation efficiency;
- The number of full-time teachers and the number of students in school have relatively high importance scores, indicating that the matching degree between teacher resources and student demand has an important impact on resource allocation efficiency;

TABLE 4. Ranking of Feature Variable Importance

Ranking	Feature Variable	Importance Score
1	Public financial investment in education (X4)	0.192
2	Number of full-time teachers (X1)	0.168
3	Number of students in school (X6)	0.141
4	Per-student value of teaching instruments and equipment (X2)	0.103
5	Per-student building area of school buildings (X3)	0.078
6	Urbanization rate (X13)	0.065
7	School enrollment rate of school-age children (X7)	0.052
8	Book collection (X5)	0.043
9	Traffic accessibility (X14)	0.037
10	Urban-rural student ratio (X9)	0.031
11	Proportion of left-behind children (X8)	0.026
12	Proportion of ethnic minority students (X10)	0.021
13	Quality of educational administrative personnel (X15)	0.017

5) Empirical Analysis of Influencing Factors on Educational Resource Allocation Efficiency

To further analyze the influence mechanism of core feature variables on educational resource allocation efficiency, this paper constructs a multiple linear regression model for empirical analysis, taking the top 5 core feature variables identified by the improved random forest algorithm as explanatory variables and the educational resource allocation efficiency value as the explained variable.

TABLE 5. Multiple Linear Regression Results

Variable	Coefficient	t-statistic
Constant term	0.235	5.682
Public financial investment in education (X4)	0.312	8.973
Number of full-time teachers (X1)	0.268	7.845
Number of students in school (X6)	-0.187	-6.234
Regional GDP (X11)	0.156	5.128
Per-student value of teaching instruments and equipment (X2)	0.123	4.357
R^2	0.896	-
Adjusted R^2	0.887	-
F-statistic	98.765	-

6) Econometric Cross-Validation

To complement the machine learning results and address small dataset limitations, a fixed-effects panel regression model was estimated using the same 155 samples. The results (Table 6) confirm the key findings from the improved random forest: public financial investment (X4), number of full-time teachers (X1), and per-student equipment value (X2) have significant positive effects on allocation efficiency, while the number of students (X6) has a significant negative effect. This cross-validation strengthens the reliability of the empirical conclusions.

TABLE 6. Fixed-Effects Panel Regression Results

Variable	Coefficient	t-statistic
Public financial investment in education (X4)	0.298	7.832
Number of full-time teachers (X1)	0.245	6.915
Number of students in school (X6)	-0.176	-5.743
Per-student value of teaching instruments and equipment (X2)	0.118	4.029
Per-student building area of school buildings (X3)	0.072	3.451
Province fixed effects	Included	-
Year fixed effects	Included	-
R^2	0.887	-
Adjusted R^2	0.876	-
F-statistic	92.47	-

VI. POLICY RECOMMENDATIONS FOR IMPROVING EDUCATIONAL RESOURCE ALLOCATION EFFICIENCY

A. OPTIMIZE THE STRUCTURE OF PUBLIC FINANCIAL INVESTMENT IN EDUCATION

Based on the empirical analysis results, public financial investment in education is a key factor affecting educational resource allocation efficiency:

(1) Increase public financial investment in education in central and western regions and rural areas to narrow the gap in educational resources between regions and urban and rural areas;

(2) Establish a dynamic adjustment mechanism for public financial investment in education, and rationally allocate financial funds according to the actual needs of each region, such as the number of students in school and teacher resources;

(3) Strengthen the supervision of the use of public financial investment in education to ensure that the funds are used for special purposes and improve the transparency and rationality of fund use.

B. STRENGTHEN THE OPTIMIZATION OF TEACHER RESOURCE ALLOCATION

Teacher resources are the core component of educational resources:

(1) Establish a teacher mobility mechanism, encourage high-quality teachers to flow to central and western regions and rural areas, and improve the quality of teachers in underdeveloped areas through teaching support, job rotation and other methods;

(2) Increase the training of teachers in rural areas and remote areas to improve their professional quality and teaching ability;

(3) Improve the teacher incentive mechanism, increase the treatment level of teachers in rural areas and remote areas, and attract more outstanding talents to engage in grassroots education.

C. PROMOTE PRECISE MATCHING OF EDUCATIONAL RESOURCE SUPPLY AND DEMAND

The matching degree between educational resource supply and demand is the core of improving resource allocation efficiency:

(1) Use technologies to establish an educational resource demand prediction platform to accurately identify the hotspots and gaps of educational resource demand in various regions;

(2) Rationally adjust the structure of educational resource supply according to the demand prediction results, and optimize school layout, teaching facility configuration, etc.;

(3) Pay attention to the educational needs of special groups, increase the inclination of educational resources to groups such as left-behind children and ethnic minority students, and ensure educational equity.

D. IMPROVE THE EXTERNAL ENVIRONMENT FOR EDUCATIONAL RESOURCE ALLOCATION

The external environment such as regional economic development level and traffic accessibility:

(1) Strengthen the economic construction of central and western regions and rural areas to enhance their capacity to carry educational resources;

(2) Strengthen the construction of infrastructure such as transportation and communication in rural areas and remote areas to improve the basic conditions for educational resource allocation;

(3) Improve the quality and management level of educational administrative personnel, and improve the scientificity and execution efficiency of educational resource allocation decisions.

VII. CONCLUSION

Focused on the specialized training needs of the industry, this research utilizes improved machine learning algorithms to enhance the efficiency of educational resource distribution, substantiating its findings through rigorous theoretical analysis and focused empirical validation:

(1) The constructed evaluation index system for educational resource allocation efficiency covers three dimensions: resource supply, resource demand, and resource allocation environment, which can comprehensively reflect the factors affecting resource allocation efficiency;

(2) By introducing an adaptive weighting mechanism and cross-validation optimization strategy, the improved random forest algorithm has better prediction accuracy and stability than the traditional random forest algorithm, and the accuracy is 12.3% higher than that of the traditional random forest algorithm;

(3) Public financial investment in education, number of full-time teachers, regional GDP, per-student teaching instruments and equipment value are significantly positively correlated with efficiency, while the number of students in school is significantly negatively correlated with educational resource allocation efficiency;

(4) Measures such as optimizing the structure of public financial investment in education, strengthening teacher resource allocation, promoting precise matching of resource supply and demand, improving the external environment.

AUTHOR CONTRIBUTIONS

Ruijun Ban designed, collected and analyzed the data, and drafted the manuscript. Ruijun Ban conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Ruijun Ban participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] G. Zhou and X. Zhou, "Education Policy and Reform in China," Singapore: Springer Nature Singapore, 2019, ch. 3, pp. 31-44, doi: 10.1007/978-981-13-6492-1_3.

[2] J. Xiao and Z. Liu, "Inequalities in the financing of compulsory education in China: A comparative study of Gansu and Jiangsu Provinces with spatial analysis," International Journal of Educational Development, vol. 39, pp. 250-263, 2014, doi: 10.1016/j.ijedudev.2014.05.004.

[3] F. Çelik and M. H. Baturay, "Technology and innovation in shaping the future of education," Smart Learning Environments, vol. 11, no. 1, pp. 54, 2024, doi: 10.1186/s40561-024-00339-0.

[4] Y. Wang, H. Lin, L. Sun, D. Zhang, J. Wang, and L. She, "Research on the Balanced Development of High Quality Compulsory Education," Proceedings of the 2020 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE); 17-19 December 2020; Frankfurt, Germany. Piscataway, NJ, USA: IEEE; 2020. pp. 262-265, doi: 10.1109/ISAIEE51769.2020.00065.

[5] Z. Ye, R. Khanal, and Z. Cao, "Studying on the efficiency of higher education resource allocation and its influencing factors in the western China using DEA-Malmquist and Tobit models," PLoS One, vol. 20, no. 10, Art. no. e0334090, 2025, doi: 10.1371/journal.pone.0334090.

[6] Q. Zhang, "Study on "hollowing out" rural governance from the perspective of precise poverty alleviation," Population, Resources & Environmental Economics, vol. 4, pp. 140-151, 2023, doi: 10.23977/pree.2023.040117.

[7] F. Li and I. Hardy, "Acculturating for Change: An Institutional Logics Perspective on Postgraduate Policy Implementation in China," Higher Education Policy, pp. 1-25, 2025, doi: 10.1057/s41307-025-00410-6.

[8] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," Neurocomputing, vol. 415, pp. 295-316, 2020, doi: 10.1016/j.neucom.2020.07.061.

[9] A. Semwal, X. Yue, Y. Shen, and M. Aibin, "Cloud Resource Allocation Recommendation Based on Machine Learning," Proceedings of the 24th International Conference on Transparent Optical Networks (ICTON 2024); 14-18 July 2024; Bari, Italy. Piscataway, NJ, USA: IEEE; 2024. pp. 1-4, doi: 10.1109/ICTON62926.2024.10647993.

[10] M. Ren, L. Yang, Y. Zhou, L. Lv, J. Chen, P. Xiao, et al., "Resource Allocation in Cell-Free MEC Networks: A Hierarchical MADRL-Based Algorithm," IEEE Transactions on Cognitive Communications and Networking, vol. 12, pp. 4593-4609, 2025, doi: 10.1109/TCCN.2025.3645433.

[11] A. Rathee and S. Dalal, "A systematic literature review of machine learning-based resource allocation techniques in cloud computing," Computing, vol. 107, no. 9, pp. 179, 2025, doi: 10.1007/s00607-025-01526-8.

[12] D. Deng, "A Performance Evaluation Model Based on AHP and Its Application," Proceedings of the 2021 International Conference of Social Computing and Digital Economy (ICSCDE); 28–29 August 2021; Chongqing, China. Piscataway, NJ, USA: IEEE; 2021. pp. 26-29, doi: 10.1109/ICSCDE54196.2021.00015.

[13] J. F. Chen, H. N. Hsieh, and Q. H. Do, "Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach," Applied Soft Computing, vol. 28, pp. 100-108, 2015, doi: 10.1016/j.asoc.2014.11.050.

[14] J. Zheng, J. Yu, L. Jin, and Z. Zhu, "An Empirical Study on the evaluation of primary compulsory education resource allocation based on TOPSIS Model with entropy weight method," Proceedings of the 2022 10th International Conference on Orange Technology (ICOT); 10-11 November 2022; Shanghai, China. Piscataway, NJ, USA: IEEE; 2022. pp. 1-4, doi: 10.1109/ICOT56925.2022.10008179.

[15] X. Zhao and Z. J. Hu, "The application of DEA model in the evaluation of higher education resource allocation— Taking Guizhou Province as an example," Educational Theory and Practice: Subject Edition, vol. 29, no. 30, pp. 15-17, 2009.

[16] Y. H. Guo and H. P. Xue, "An analysis of the evaluation of the allocation efficiency of primary and secondary school educational resources based on DEA method—Taking Hubei and Jiangsu provinces in central and eastern China as examples," Proceedings of the 2009 Academic Annual Conference of China Educational Economics Association; 5th December 2009; Guangzhou, Guangdong, China. Guangzhou, Guangdong, China: Educational Economics Branch of the Chinese Society of Education; 2009. pp. 235-255.

[17] S. Y. Zou, F. M. Li, A. P. Xie, and T. Zhou, "A resource allocation method based on random forest algorithm: China, 2020113133348.8," 26th February 2021.

[18] X. Z. Qian, J. Qin, and W. Song, "Improved parallel random forest algorithm and its out-of-bag estimation," Application Research of Computers, vol. 35, no. 6, pp. 4, 2018, doi: 10.3969/j.issn.1001-3695.2018.06.011.

[19] K. N. Fang, J. B. Wu, J. P. Zhu, and B. C. Xie, "A review of random forest method research," Statistics & Information Forum, vol. 26, no. 3, pp. 7, 2011, doi: 10.3969/j.issn.1007-3116.2011.03.006.

[20] K. Jurczuk, M. Czajkowski, and M. Kretowski, "Understanding evolutionary induction of decision trees," Proceedings of the Genetic and Evolutionary Computation Conference Companion; 10-14 July 2021; Lille, France. New York, USA: Association for Computing Machinery; 2021. pp. 155-156, doi: 10.1145/3449726.3459422.

[21] M. Krzywinski and N. Altman, "Classification and regression trees," Nature Methods, vol. 14, no. 8, pp. 757-758, 2017, doi: 10.1038/nmeth.4370.