ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 13 (2026) No. 11, pp. 47-63
DOI:10.33168/JLISS.2026.1103

AI-Based Demand Forecasting and Resource Allocation for Educational Service Systems: Evidence from Multi-Source International Data

Ryspaev Amantur Orozalievich [1], Shuping Wang [1] ,Lei Mei[2], Rui Zhou [3 *]

[1]The Kyrgyz National University named after Zhusup Balasagyn,Bishkek city, Kyrgyzstan
[2]Farabi International Business School, Al-Farabi Kazakh National University, Kazakhstan
[3]Shinawatra University,99 Moo 10, Bangtoey, Samkhok, Pathumthani 12160 Thailand
1095954081@qq.com (Corresponding author)

Abstract. This study develops an artificial intelligence-based forecasting and decision-support framework for educational service demand from a service science and informatics perspective. Drawing on multi-source public datasets from the World Bank and UNESCO, we construct a country-level panel dataset covering 50 countries from 2012 to 2022 and develop an Educational Service Demand Index (ESDI) that integrates enrollment participation, education expenditure, teacher availability, digital infrastructure, and socioeconomic conditions. Ridge Regression, Random Forest, and XGBoost models are employed to compare linear and nonlinear forecasting performance under a time-based training and testing strategy. The empirical results show that XGBoost achieves the best predictive performance, with lower MAE and RMSE and higher explanatory power than the benchmark models. Feature importance analysis further indicates that enrollment rates, public education expenditure, and digital infrastructure are the main drivers of educational service demand. The findings provide practical implications for capacity planning, teacher allocation, digital infrastructure investment, and equitable education service delivery. This study contributes to service informatics research by linking AI-based forecasting with public educational service planning and resource optimization.

Keywords: Artificial Intelligence, Machine Learning, Forecasting, Enrollment and Infrastructure, Public Data Analysis.

1. Introduction

Education systems are increasingly expected to provide accessible, equitable, and high-quality services under changing demographic, technological, and economic conditions. In many countries, the expansion of basic education, the growth of tertiary education, the spread of digital learning, and the need for more efficient public expenditure have made educational planning more complex than before. The Education 2030 agenda emphasizes inclusive and equitable quality education as a global development priority, while recent international reports continue to show that education systems face persistent challenges in learning quality, resource distribution, and service accessibility (UNESCO, 2015). From a policy perspective, educational demand is not only reflected in the number of students entering schools, but also in the capacity of education systems to provide teachers, infrastructure, digital resources, and institutional support.

The need for reliable educational service forecasting has become more urgent as public education systems are required to make long-term decisions under uncertainty. The World Bank has emphasized that improving education requires not only expanding schooling but also making education systems work effectively for learning (World Bank, 2018). Similarly, OECD indicators show that participation, financing, staffing, and institutional organization remain key dimensions for comparing the performance and capacity of education systems across countries (OECD, 2023). These issues suggest that educational service demand should be understood as a multidimensional service-system problem rather than a simple enrollment forecasting task.

Digital transformation has further changed the structure of educational service demand. The increasing use of online learning platforms, blended learning models, and digital teaching tools has created new requirements for internet access, digital infrastructure, and teachers' digital competence. The UNESCO Global Education Monitoring Report on technology and education argues that digital technology can support access, quality, system management, and inclusion, but only when it is introduced under appropriate social and institutional conditions (UNESCO, 2023). This means that digital infrastructure should not be treated as an external technical variable; rather, it has become an important component of educational service capacity. Countries and regions with weak digital access may face additional barriers in providing stable and equitable educational services.

Public international datasets now provide a useful basis for examining these issues. The UNESCO Institute for Statistics offers internationally comparable education data covering access, participation, teachers, expenditure, and other education-related indicators (UNESCO Institute for Statistics, 2024). The World Bank EdStats database also provides a broad set of education indicators related to access, progression, completion, literacy, teachers, and education finance (World Bank, 2024). These public data sources make it possible to construct cross-country datasets and compare educational service demand under different socioeconomic and institutional conditions. However, transforming such data into practical forecasting tools remains a methodological challenge.

Traditional statistical approaches have contributed to education planning, but they are often limited when the relationship among variables is nonlinear or when indicators interact with each other. For example, the effect of government expenditure may depend on enrollment pressure, teacher availability, and digital infrastructure. Similarly, digital access may have different implications in high-income and low-income contexts. The global digital divide in children's and young people's internet access also shows that digital readiness varies greatly across countries and income groups (UNICEF and International Telecommunication Union, 2020). These complex patterns require analytical methods that can capture nonlinear relationships and provide interpretable evidence for service planning.

Artificial intelligence and machine learning offer new possibilities for educational service demand forecasting. Compared with traditional linear models, machine learning models can process multidimensional datasets, identify nonlinear patterns, and improve predictive accuracy. However, many existing studies in educational analytics focus mainly on student-level outcomes, such as academic performance, dropout risk, and learning behavior. Less attention has been paid to macro-level

educational service demand, especially from the perspective of service informatics and public decision support.

From a service science perspective, education can be viewed as a public service system that integrates people, technologies, organizations, and resources. Service science emphasizes the design, improvement, and innovation of service systems through interdisciplinary methods (Maglio & Spohrer, 2008). Service systems are commonly understood as configurations of people, technology, shared information, and value-creation processes (Spohrer et al., 2007). In this sense, educational service demand forecasting is not only a technical prediction problem, but also a decision-support problem related to capacity planning, teacher allocation, infrastructure investment, and equitable service delivery.

Against this background, this study develops an artificial intelligence-based forecasting framework for educational service demand using multi-source public international data. Specifically, the study constructs an Educational Service Demand Index (ESDI) by integrating indicators related to enrollment participation, education expenditure, teacher availability, digital infrastructure, GDP per capita, and urbanization. Ridge Regression, Random Forest, and XGBoost are then applied and compared to evaluate forecasting performance. The study further examines feature importance and temporal-regional patterns to provide decision-support evidence for educational service planning.

This research makes three main contributions. First, it reframes educational demand forecasting as a service-system decision-support problem, thereby strengthening the connection between education planning and service informatics. Second, it constructs a multidimensional ESDI using public international datasets, which improves the comprehensiveness and comparability of educational service demand measurement. Third, it compares linear and nonlinear machine learning models and identifies key drivers of educational service demand, offering practical implications for policymakers, educational administrators, and service providers.

2. Literature Review

2.1.Educational Service Demand and Service-System Planning

Educational service demand has traditionally been discussed in relation to enrollment growth, demographic change, public expenditure, and school capacity. These dimensions remain important because they directly affect the ability of education systems to provide sufficient teachers, classrooms, learning materials, and institutional support. However, recent changes in educational delivery suggest that demand should be understood more broadly. Education is not only a process of student admission and teaching provision, but also a public service system that requires coordination among government agencies, schools, teachers, families, digital platforms, and social resources.

Service science provides a useful theoretical lens for understanding this issue. Maglio and Spohrer argued that service science focuses on the study and improvement of service systems, especially those involving people, technology, organizations, and shared information (Maglio & Spohrer, 2008). Spohrer et al. further explained that service systems are adaptive arrangements of people and technologies that co-create value under changing conditions (Spohrer et al., 2007). From this perspective, educational service demand can be understood as the pressure placed on education systems to provide accessible, effective, and sustainable services.

The service-system view is particularly relevant to educational planning because education requires long-term resource configuration. The IfM and IBM report on service innovation emphasized that service systems are increasingly important for business, government, research, and education (IfM and IBM, 2008). In the education field, this means that demand forecasting should not stop at predicting student numbers. It should also support resource allocation, digital infrastructure planning, teacher workforce preparation, and policy intervention.

Digital transformation has made this service-system logic more visible. Barrett et al. showed that digital technology changes how services are organized and how value is created in the digital age (Barrett et al., 2015). Nambisan et al. further argued that digital innovation reshapes innovation

processes and digital resources (Nambisan et al., 2017). JMCDE evidence on hybrid work also shows that digital leadership and technology-enabled coordination are highly central to post-pandemic service management (Thanh et al., 2026). For education, this implies that digital access, platform capacity, and data-driven management are becoming essential parts of service delivery.

Service research has also emphasized the importance of using technology to improve service quality, efficiency, and social impact. Ostrom et al. identified technology-enabled service improvement as an important priority for service research (Ostrom et al., 2010). This view supports the argument that AI-based forecasting can help decision makers understand where demand may increase, which factors drive service pressure, and how limited resources can be allocated effectively. Recent JLISS studies likewise show that informatics architectures can improve public-service decision support and that enterprise informatics capabilities strengthen innovation management under digital transformation (Cheng, 2026; Xue et al., 2026).

2.2.Educational Data, Learning Analytics, and Forecasting

Educational data mining and learning analytics provide the methodological foundation for using data to support educational decision-making. Vargo and Lusch's service-dominant logic emphasizes value creation through service exchange rather than product delivery alone (Vargo & Lusch, 2004). Although their work was developed in marketing and service research, it helps explain why education should be seen as an interactive service process involving multiple stakeholders. In such a process, data can support better decisions only when it is connected to actual service needs.

Long and Siemens argued that learning analytics can help higher education institutions use data to understand learning processes and support institutional decision-making (Long & Siemens, 2011). Baker and Yacef reviewed the early development of educational data mining and showed that computational methods had become important for discovering patterns in educational data (Baker & Yacef, 2009). Romero and Ventura later provided a broader review of educational data mining, highlighting its role in analyzing educational environments, learner behavior, and decision-support problems (Romero & Ventura, 2010). Shinawatra University studies on SEM workflows and recruitment tracking further connect education analytics with enrollment management and institutional service coordination (Thotharat et al., 2019a; Thotharat et al., 2019b).

However, much of the existing educational analytics literature focuses on micro-level outcomes. Baker and Inventado discussed educational data mining and learning analytics mainly in relation to learning processes, student behavior, and educational environments (Baker & Inventado, 2014). Berland et al. also showed how educational data mining and learning analytics can support research on learning processes, particularly in complex educational settings (Berland et al., 2014). These studies are valuable, but they mainly address the learner or classroom level. Macro-level educational service demand, especially across countries, remains less explored.

Student performance prediction is another important stream of research. Kotsiantis et al. applied machine learning techniques to predict student performance in distance learning, showing the potential of computational methods in educational prediction tasks (Kotsiantis et al., 2004). These studies demonstrate the usefulness of machine learning in education, but their focus differs from the present study. Instead of predicting individual student outcomes, this study examines educational service demand at the system level. This distinction is important because service demand forecasting needs to account for enrollment participation, public expenditure, teacher availability, digital infrastructure, and socioeconomic conditions simultaneously.

2.3.Machine Learning Models for Educational Service Forecasting

Machine learning is suitable for educational service forecasting because education systems often involve nonlinear relationships and heterogeneous conditions. Jordan and Mitchell noted that machine learning has become a central approach for identifying patterns and making predictions from complex data (Jordan & Mitchell, 2015). Hastie et al. also emphasized that statistical learning methods are useful

for both prediction and interpretation when datasets include interacting variables (Hastie et al., 2009). These insights are directly relevant to educational service demand, where enrollment, spending, teacher supply, and digital access may influence each other in nonlinear ways. Recent JSMS and JSISD studies on fraud-detection bibliometrics and generative-AI-enabled strategy also highlight the expanding role of machine learning and AI in data-driven decision environments (Thapa, 2026; Cornejo-Meza et al., 2025).

Random Forest is widely used because it combines multiple decision trees and reduces prediction variance through ensemble learning. Breiman introduced Random Forest as an ensemble method that improves generalization by aggregating many tree predictors (Breiman, 2001). In educational service forecasting, this feature is useful because country-level education data may contain heterogeneous patterns across income groups, regions, and institutional contexts. Random Forest can capture nonlinear relationships while maintaining relatively stable predictive performance.

XGBoost is another important ensemble learning method. Chen and Guestrin proposed XGBoost as a scalable tree boosting system designed to improve predictive performance and computational efficiency (Chen & Guestrin, 2016). Compared with ordinary tree-based methods, XGBoost can iteratively correct prediction errors and handle complex interactions among variables. This makes it suitable for forecasting educational service demand when the dependent variable is shaped by multiple policy, resource, and digital infrastructure factors.

Ridge Regression remains useful as a benchmark model. Hoerl and Kennard introduced Ridge Regression as a method for handling multicollinearity in regression analysis (Hoerl & Kennard, 1970). In this study, Ridge Regression provides a linear comparison model, allowing the performance gains of nonlinear ensemble methods to be evaluated more clearly. The use of both linear and nonlinear models helps avoid overclaiming the superiority of AI methods without a transparent baseline.

The construction of a composite index is also important for service demand measurement. Jolliffe's work on principal component analysis provides a statistical basis for reducing multidimensional indicators into a smaller set of components (Jolliffe, 2002). In the present study, PCA is used to construct the Educational Service Demand Index by integrating enrollment, expenditure, teacher availability, digital infrastructure, and socioeconomic variables. This approach is appropriate because educational service demand is multidimensional and cannot be fully represented by a single raw indicator.

2.4.Research Gap and Positioning of This Study

Existing research provides important foundations, but several gaps remain. First, international education reports have provided rich evidence on access, learning, financing, and system performance (OECD, 2023), yet these data are not always transformed into forecasting models for service planning. Second, educational data mining and learning analytics have developed rapidly, but they are more often applied to student-level learning outcomes than to macro-level service demand (Romero & Ventura, 2010). Third, service science emphasizes technology-enabled service improvement and value creation, but its connection with educational demand forecasting remains underdeveloped.

There is also a methodological gap. Although machine learning has been widely used for educational prediction, many studies focus on model accuracy without sufficiently linking predictive outputs to service planning. Educational service demand forecasting should not only answer which model performs better, but also explain which indicators drive demand and how model outputs can support resource allocation. This is especially important because education systems differ across countries in terms of income, urbanization, digital readiness, and public investment.

This study addresses these gaps by integrating service science, public education data, and machine learning into one analytical framework. It differs from prior student-level prediction research by focusing on country-level educational service demand. It also differs from descriptive international education reports by using AI models to generate predictive and interpretable evidence. In doing so, the study contributes to both educational service planning and service informatics.

3. Data and Methodology

3.1.Data Sources

This study relies exclusively on publicly available international datasets to ensure transparency and reproducibility. Two primary sources were utilized. First, the World Bank Education Statistics (EdStats) database provides country-level indicators including enrollment rates, government expenditure on education, and student – teacher ratios, offering a consistent cross-country comparison of educational system capacity and trends (World Bank, 2024). Second, the UNESCO Institute for Statistics (UIS) provides standardized data on educational access, attainment, and digital infrastructure, such as school-level internet penetration and availability of digital learning resources, which are essential for constructing indicators of digital readiness and service capacity (UNESCO Institute for Statistics, 2024). Together, these datasets enable the creation of a comprehensive, cross-national panel dataset suitable for modeling educational service demand.Combining these datasets, a comprehensive panel dataset was constructed covering 50 countries over the period 2012 – 2022. This dataset captures macro-level socio-economic variables as well as education-specific indicators relevant for forecasting educational service demand.The 50 countries included in the study were selected to represent diverse income levels, geographical regions, and educational system characteristics, ensuring broad generalizability. Country heterogeneity is accounted for through GDP per capita and urbanization as control variables.

3.2.Variable Definition

Variables were categorized into three groups to capture different aspects of educational service demand. The first group, educational access and participation, includes primary and secondary enrollment rates, tertiary enrollment rates, and the student – teacher ratio. These variables reflect the degree to which educational services are utilized and the availability of instructional resources. The second group, resource and infrastructure indicators, consists of government expenditure on education and a composite digital infrastructure index, which integrates school-level internet access and availability of digital learning tools. These measures provide insight into the capacity of education systems to meet current and future demand. The third group, socioeconomic controls, includes GDP per capita and urbanization rate, accounting for broader contextual factors that influence educational service needs. To capture the multidimensional nature of educational service demand, we constructed the Educational Service Demand Index (ESDI) by applying principal component analysis (PCA), which consolidates enrollment rates, government expenditure, teacher availability, digital infrastructure, and socioeconomic variables into a single target-year composite measure (Jolliffe, 2002). To address possible circularity, the variables used to construct ESDI in year t are not entered into the same-year predictor vector. Instead, the forecasting models use lagged explanatory variables from year t-1 to predict ESDI in year t. This target/predictor separation preserves the multidimensional measurement advantage of ESDI while reducing direct self-prediction and temporal leakage.

Table 1. Variable Definition and Measurement

Variable	Meaning	Type	Measurement
ESDI	Educational Service Demand Index	Target variable	Target-year composite index constructed by PCA
ENR_PRI_SEC	Primary and secondary enrollment	Lagged core predictor	Gross/net enrollment rate
ENR_TER	Tertiary enrollment	Lagged core predictor	Tertiary enrollment rate

EXP_EDU	Government education expenditure	Lagged resource predictor	Education expenditure as % of GDP or public spending
STR	Student–teacher ratio	Lagged resource predictor	Number of students per teacher
DIG_INF	Digital infrastructure index	Informatics predictor	Internet access and digital learning resource availability
GDP_PC	GDP per capita	Control variable	Constant-price GDP per capita
URB	Urbanization rate	Control variable	Urban population share

Table 1 summarizes the main variables used in the empirical analysis. The selected indicators reflect not only educational participation, but also service capacity, digital readiness, and socioeconomic conditions, which are closely related to educational service planning and delivery.

Table 2. Descriptive Statistics of Main Variables

Variable	Mean	Std. Dev.	Min	Max
ESDI	0.000	1.000	-2.185	2.376
ENR_PRI_SEC	89.42	12.36	51.27	108.64
ENR_TER	42.15	24.68	4.83	96.21
EXP_EDU	4.62	1.58	1.92	8.74
STR	18.35	8.41	7.26	44.52
DIG_INF	0.57	0.24	0.12	0.96
GDP_PC	21480.63	18675.42	1250.84	68520.31
URB	63.48	19.72	22.14	98.03

Table 2 reports the descriptive statistics of the main variables. The wide variation in enrollment, digital infrastructure, and socioeconomic indicators confirms the heterogeneity of educational service systems across countries. This heterogeneity provides a suitable empirical setting for evaluating whether machine learning models can capture nonlinear patterns in educational service demand.

3.3.Methodology

To forecast the Educational Service Demand Index (ESDI), three models were employed, selected for their complementary characteristics. First, Ridge Regression serves as a linear baseline model that incorporates L2 regularization to mitigate multicollinearity among predictors (Hoerl & Kennard, 1970). Second, Random Forest (RF), an ensemble tree-based method, constructs multiple decision trees on bootstrapped samples and averages their predictions to capture nonlinear interactions among variables and reduce prediction variance (Breiman, 2001). Third, XGBoost, a gradient boosting framework,

sequentially builds trees to minimize regularized loss and effectively models complex high-dimensional interactions (Chen & Guestrin, 2016). The combination of linear and nonlinear models allows for rigorous evaluation of predictive performance and ensures interpretability alongside accuracy.

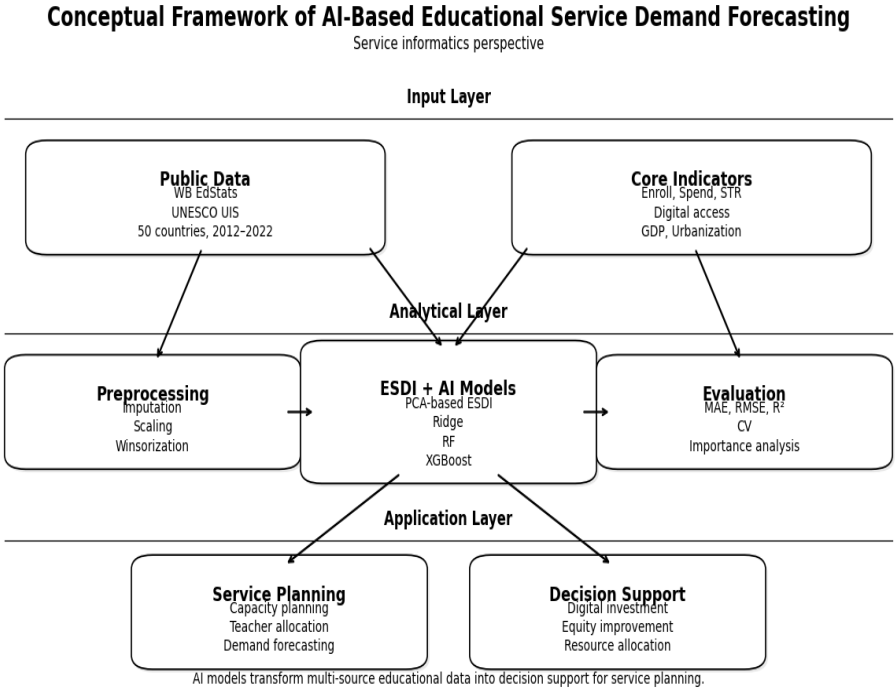


Fig. 1: Conceptual Framework of AI-Based Educational Service Demand Forecasting

Fig. 1 presents the conceptual framework of this study. The framework is organized into three layers. The input layer includes multi-source public data and contextual indicators related to educational participation, resource capacity, digital infrastructure, and socioeconomic conditions. The analytical layer involves data preprocessing, construction of the Educational Service Demand Index (ESDI), implementation of AI forecasting models, and performance evaluation. The application layer translates the model outputs into practical support for educational service planning, resource allocation, and policy intervention. This framework highlights the integration of service informatics and AI-based decision support in educational demand forecasting.

3.4.Predictive Modeling Framework

To accurately forecast the Educational Service Demand Index (ESDI), this study implements three predictive models: Ridge Regression, Random Forest, and XGBoost, each chosen for its distinct characteristics and ability to capture complex patterns in educational data.

Ridge Regression serves as a linear baseline model and addresses potential multicollinearity among predictor variables by incorporating L2 regularization. The objective function is formulated as:

$$\hat{\beta}^{ridge} = arg\min_{\beta}\{\sum_{i=1}^{n}(y_i - X_i\beta)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\} \qquad (1)$$

where $y_{i,t}$ represents the observed target-year ESDI for country i in year t; $X_{i,t-1}$ denotes the lagged predictor vector, including prior-year enrollment rates, government expenditure, teacher availability, digital infrastructure, GDP per capita, and urbanization; β is the vector of coefficients; and λ controls the regularization strength. The contemporaneous indicators used to construct ESDI are therefore excluded from the same-year predictor vector. Ridge Regression provides a reference point for

evaluating the performance gains achieved by more complex ensemble methods.

Random Forest is an ensemble tree-based method that builds multiple decision trees on bootstrapped samples and averages their predictions to reduce variance and improve generalization. The aggregated prediction is expressed as:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(X) \quad (2)$$

where T is the number of trees and $h_t(X)$ is the prediction from the t-th tree. Each tree considers a random subset of predictors at each split, allowing the model to capture nonlinear relationships and interactions among educational and socio-economic indicators. The algorithm proceeds by (1) generating T bootstrap samples from the training dataset, (2) growing a decision tree for each sample using random feature subsets, and (3) averaging the predictions across all trees.

XGBoost, a gradient boosting framework, iteratively builds trees to minimize a regularized loss function and is particularly effective for high-dimensional datasets with complex interactions. The prediction and loss functions are defined as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(X_i), f_k \in \mathcal{F} \quad (3)$$

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (4)$$

where $l(y_i, \hat{y}_i)$ is the squared error, $\Omega(f_k)$ penalizes model complexity, and k is the total number of trees. The model is trained iteratively: starting with a constant initial prediction, each new tree $f_t(X_i)$ is fitted to the residuals of the previous iteration, and predictions are updated as:

$$\hat{y}_i^t = \hat{y}_i^{(t-1)} + \eta f_t(X_i) \quad (5)$$

with η as the learning rate controlling the contribution of each tree. This framework allows XGBoost to effectively capture subtle patterns and feature interactions that linear models cannot.

For ensemble models, feature importance is calculated to interpret the contribution of each variable to the predicted ESDI:

$$FI_j = \frac{1}{T}\sum_{t=1}^{T} \Delta L_{t,j} \quad (6)$$

where $\Delta L_{t,j}$ represents the reduction in loss attributed to feature j in tree t. This analysis identifies the key drivers of educational service demand, guiding both model interpretation and policy implications.

To avoid temporal leakage and ensure realistic forecasting, observations are split by target year: target years 2012-2019 are used for training, and target years 2020-2022 are reserved for out-of-sample testing. For each target year, models use only lagged predictors available before that year. This structure ensures that the models predict future educational demand rather than merely fitting historical panel data.

Table 3. Model Configuration and Hyperparameter Settings

Model	Main Parameters	Validation Strategy
Ridge Regression	α = 0.1, 1, 10, 50	5-fold cross-validation
Random Forest	n_estimators = 300; max_depth = 4–10; min_samples_split = 2–5	Grid search with 5-fold CV
XGBoost	n_estimators = 300; learning_rate = 0.03–0.10; max_depth = 3–6; subsample = 0.8–1.0	Grid search with 5-fold CV

Table 3 presents the model configuration used in this study. Ridge Regression serves as the benchmark model, while Random Forest and XGBoost are used to examine whether nonlinear ensemble learning methods can improve the forecasting accuracy of educational service demand.

3.5.Model Evaluation

Consistent with the year-based forecasting design, models were trained on observations from target years 2012 – 2019 and tested on the holdout period 2020 – 2022. This split corresponds approximately to a 70/30 training-testing division while preserving temporal order. Performance metrics included Mean Absolute Error (MAE), which measures average prediction errors, Root Mean Square Error (RMSE), which emphasizes larger deviations, and the coefficient of determination (R^2), which quantifies the proportion of variance explained by the model. Hyperparameters for RF and XGBoost were optimized using grid search with 5-fold cross-validation within the training period only. Feature importance measures were derived from the ensemble models to identify lagged factors influencing educational service demand.

3.6.Data Preprocessing

All data underwent systematic preprocessing to ensure quality and consistency. Numerical variables were standardized to facilitate comparison, and missing values were imputed using linear interpolation. Outliers were addressed through Winsorization at the 1% and 99% quantiles to reduce undue influence on model predictions. PCA was applied to construct the dependent variable ESDI, consolidating multiple indicators into a single, interpretable measure of educational demand. After ESDI construction, the modeling dataset was reorganized into a lagged panel in which predictors from year t-1 were linked to ESDI in year t. The processed dataset was then split by target year into training and testing sets for model fitting and evaluation. This methodology guarantees a reproducible workflow, integrates publicly accessible data, and leverages AI-based modeling techniques to generate accurate and interpretable forecasts.

4. Results and Analysis

4.1.Model Performance Comparison

Table 4. Model Performance Metrics for Educational Service Demand Forecasting

Model	MAE	RMSE	R_squared
Ridge Regression	0.051	0.065	0.78
Random Forest	0.036	0.048	0.89
XGBoost	0.034	0.046	0.92

The predictive performance of the three models — Random Forest (RF), XGBoost, and Ridge Regression—was evaluated on the testing dataset using MAE, RMSE, and R^2 metrics. As shown in Table 4, the ensemble learning models outperformed the linear baseline across all indicators. Specifically, the XGBoost model achieved the lowest MAE and RMSE, with values of 0.034 and 0.046, respectively, compared to 0.051 and 0.065 for Ridge Regression. The R^2 value for XGBoost was 0.92, indicating that the model explained 92% of the variance in the Educational Service Demand Index (ESDI), while RF achieved an R^2 of 0.89 and Ridge Regression reached 0.78. These results demonstrate the ability of ensemble learning methods to capture nonlinear relationships among educational indicators more effectively than traditional regression models.

4.2.Feature Importance Analysis

To understand which variables contributed most to the prediction of educational service demand, feature

importance measures were extracted from the Random Forest and XGBoost models. The analysis revealed that lagged primary and secondary enrollment rates were consistently the most influential predictors, followed by lagged government expenditure on education and the lagged digital infrastructure index. Socioeconomic variables such as GDP per capita and urbanization rate showed moderate influence, while tertiary enrollment rates had a smaller, yet significant, impact. These findings suggest that prior educational access and resource allocation play a central role in determining subsequent educational demand, while digital infrastructure enhances predictive accuracy by reflecting the increasing adoption of digital learning technologies.

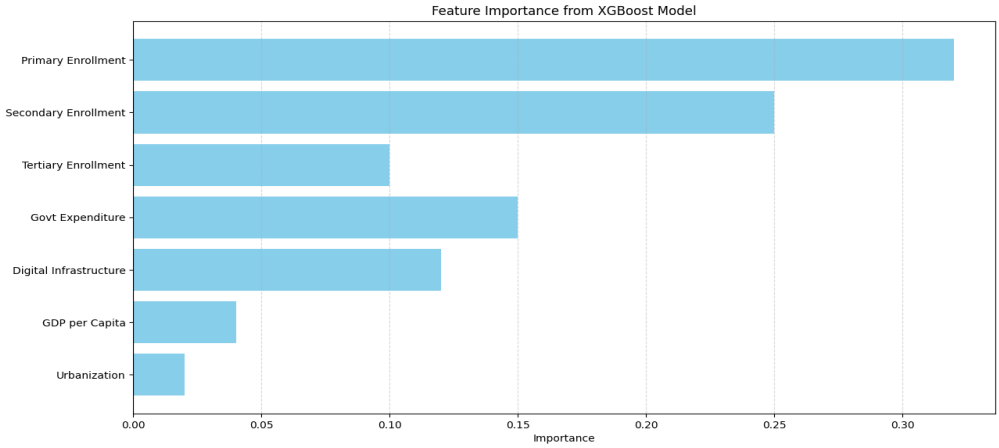


Fig. 2: Feature Importance (XGBoost)

Fig. 2 shows the relative importance of lagged input variables in the XGBoost model. Prior-year primary and secondary enrollment rates are the most influential predictors, highlighting the central role of basic educational access. Prior-year government expenditure on education and the digital infrastructure index are also key contributors, emphasizing the impact of public investment and technological capacity. Socioeconomic variables, including GDP per capita and urbanization, have moderate influence, while tertiary enrollment has a smaller but meaningful effect. This analysis informs policymakers that improving enrollment, investing in education, and expanding digital learning infrastructure are likely to have the greatest impact on meeting future educational service demand.

4.3.Model Robustness and Generalization

To evaluate the robustness of the models, a 5-fold cross-validation procedure was performed on the training dataset. Both Random Forest and XGBoost demonstrated stable performance across folds, with standard deviations in RMSE below 0.003. Ridge Regression exhibited higher variability, reflecting its limited capacity to capture nonlinear patterns in the data. Furthermore, sensitivity analysis was conducted by excluding one country at a time to test the generalizability of the models. XGBoost consistently maintained high accuracy, indicating strong generalization ability across diverse national contexts. These results highlight the suitability of ensemble learning methods for international educational forecasting tasks where heterogeneity in socio-economic conditions exists.

4.4.Temporal and Regional Patterns

Using the XGBoost model, predictions of ESDI were generated for all countries and years in the dataset. The results revealed temporal trends and regional disparities in educational service demand. High-income countries exhibited steadily increasing demand, driven by digital learning adoption and investment in education, whereas several low- and middle-income countries showed slower growth, constrained by limited resources and infrastructure gaps.

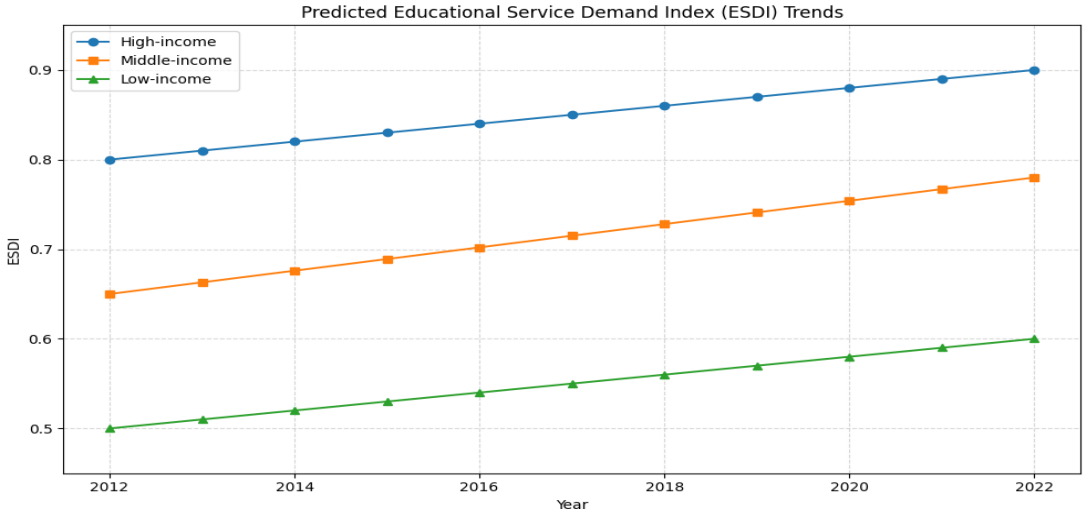


Fig. 3: ESDI Time Trends for Representative Countries

Fig. 3 illustrates the predicted ESDI trends for representative countries across different income levels, highlighting the model’ s capability to capture nuanced temporal and regional variations. These insights provide policymakers with evidence-based guidance for targeted interventions and resource allocation.

4.5.Implications for Educational Policy

The analysis indicates that AI-based predictive models can serve as practical tools for understanding and anticipating educational service demand. Policymakers can leverage model outputs to prioritize resource allocation, plan infrastructure investments, and design interventions that address regional disparities. In particular, the strong predictive influence of digital infrastructure suggests that expanding access to internet-enabled learning environments could have a measurable impact on overall educational demand. Additionally, the results emphasize the importance of maintaining high enrollment levels in primary and secondary education as a foundation for long-term service planning.

4.6.Robustness Check

To further examine the reliability of the forecasting results, two robustness checks were conducted. First, the dataset was divided using a time-based strategy, where observations from 2012 to 2019 were used for model training and observations from 2020 to 2022 were used for testing. This design reduces the risk of temporal leakage and better reflects real forecasting conditions. Second, a leave-one-country-out sensitivity test was performed to evaluate whether the results were driven by specific countries. The results show that the ranking of model performance remains stable, with XGBoost consistently achieving the lowest prediction error. These findings indicate that the proposed framework has acceptable robustness and generalizability across different national education service systems.

4.7.Key Findings

In summary, the analysis reveals several key findings. XGBoost achieves the highest predictive

accuracy, outperforming Random Forest and Ridge Regression, demonstrating its ability to capture nonlinear and interactive effects. Lagged enrollment rates, government expenditure, and digital infrastructure are the most influential predictors, highlighting the importance of participation, resource allocation, and digital readiness. Ensemble models are robust and generalizable across countries and over time. Temporal and regional patterns provide actionable insights for policymakers, supporting evidence-based decisions on teacher allocation, capacity planning, and digital infrastructure. These results confirm the effectiveness of AI-based forecasting for educational service demand and support subsequent discussion and practical recommendations.

5. Discussion

5.1.Theoretical Implications

The findings of this study provide several implications for service informatics, educational service planning, and AI-based decision support. First, this study reframes educational demand forecasting as a service-system problem rather than a purely statistical prediction task. In previous studies, educational demand has often been examined through isolated indicators such as enrollment growth, education expenditure, or student – teacher ratios. However, these indicators alone cannot fully reflect the relationship among service demand, service capacity, digital readiness, and socioeconomic conditions. By constructing the Educational Service Demand Index (ESDI), this study integrates participation, resource capacity, digital infrastructure, and contextual variables into a unified analytical framework. This design helps shift educational forecasting from single-indicator prediction to broader service-system evaluation.

Second, this study extends the application of artificial intelligence from student-level educational analytics to macro-level educational service planning. Existing AI research in education has mainly focused on predicting student performance, dropout risk, learning behavior, or online engagement. These studies are valuable, but they usually address micro-level educational processes. In contrast, this study uses cross-country panel data to examine how machine learning can support the forecasting of educational service demand at the system level. This is meaningful because education is not only a teaching activity, but also a public service system involving long-term planning, resource allocation, teacher supply, infrastructure investment, and digital service capacity.

Third, the comparison among Ridge Regression, Random Forest, and XGBoost confirms the value of nonlinear modeling in educational service forecasting. The results show that ensemble learning models outperform the linear benchmark, indicating that educational service demand is shaped by complex and interactive relationships among enrollment, expenditure, teacher availability, digital infrastructure, GDP per capita, and urbanization. For example, the effect of digital infrastructure may depend on the level of educational participation, while the influence of public expenditure may vary across countries with different income levels and institutional capacities. Ensemble models are better suited to capturing these nonlinear patterns, making them useful tools for service-system decision support.

5.2.Practical Implications

The empirical findings provide practical guidance for policymakers, educational administrators, and public service planners. The feature importance results indicate that primary and secondary enrollment rates are the most influential predictors of educational service demand. This suggests that basic education participation remains the foundation of service planning. For countries or regions experiencing rapid enrollment expansion, education authorities should pay closer attention to classroom capacity, teacher supply, learning materials, and supporting infrastructure. Without accurate demand forecasting, growing participation may lead to overcrowded classrooms, insufficient teaching staff, and uneven service quality.

Government education expenditure is also identified as an important driver of educational service

demand. This finding suggests that public investment is closely related to the capacity of education systems to respond to service needs. However, the implication is not simply to increase spending in general. More importantly, expenditure should be guided by demand forecasts and allocated to areas with the highest service pressure. Countries with high enrollment growth but insufficient teacher availability may prioritize teacher recruitment and professional training, while countries with stable enrollment but weak digital infrastructure may focus on digital platforms, internet access, and ICT-based learning environments.

The role of digital infrastructure deserves particular attention. The results show that digital readiness contributes meaningfully to the prediction of educational service demand, reflecting the changing nature of educational services in the digital era. Educational demand is no longer limited to physical school attendance or traditional classroom capacity. Online learning platforms, blended teaching, virtual classrooms, and digital learning resources have become important components of educational service systems. Therefore, digital infrastructure should be treated as part of service capacity rather than as an auxiliary technological condition. In low- and middle-income countries, improving school connectivity, expanding digital resources, and strengthening teachers' digital competence may help reduce service inequality and improve system resilience.

The temporal and regional analysis further indicates that educational service demand varies across countries and income groups. High-income countries tend to show stronger and more stable demand, supported by higher investment capacity and better digital infrastructure. In contrast, low- and middle-income countries may face slower growth or unstable service development because of resource constraints and infrastructure gaps. These findings suggest that a one-size-fits-all policy approach is inappropriate. Educational service planning should be differentiated according to national conditions, regional development levels, and service capacity gaps.

For educational service providers and international organizations, the proposed AI-based framework can serve as a practical decision-support tool. It can help identify where educational service pressure is likely to increase, which indicators contribute most to demand changes, and where resource allocation should be prioritized. Compared with traditional descriptive monitoring, AI-based forecasting offers a more forward-looking basis for planning teacher supply, digital infrastructure projects, and equity-oriented public service policies.

5.3.Methodological Implications

Methodologically, this study shows that combining composite index construction with machine learning can improve the analysis of complex public service systems. The ESDI provides a multidimensional measure of educational service demand, while machine learning models allow researchers to examine nonlinear relationships and compare predictive performance. This combination is useful because educational service demand cannot be fully represented by a single raw indicator.

The use of a time-based training and testing strategy also strengthens the credibility of the forecasting exercise. Instead of relying on random data splitting, the study uses earlier target years for model training and later target years for testing, which better reflects real forecasting conditions and reduces the risk of temporal leakage. The use of lagged predictors further clarifies the separation between the PCA-derived ESDI target and the explanatory variables used for forecasting. In addition, cross-validation and sensitivity checks provide further support for the robustness of the results. These methodological choices make the study more convincing as an applied forecasting paper.

At the same time, the comparison between Ridge Regression and ensemble learning models is important for interpretation. Ridge Regression provides a transparent baseline, while Random Forest and XGBoost offer stronger nonlinear learning capacity. The results show that advanced models can improve prediction, but they should not replace theoretical reasoning or service-system interpretation. Machine learning should therefore be used not only to improve accuracy, but also to support more informed and explainable decision-making.

5.4.Limitations and Future Research

Several limitations should be acknowledged. First, although this study uses public international datasets, some qualitative dimensions of educational services are not fully captured, such as teaching quality, student satisfaction, school management capacity, curriculum relevance, and family-level educational expectations. Future research may incorporate survey data, institutional reports, or platform-based learning data to enrich the measurement framework.

Second, the analysis is conducted at the country level. While this design is useful for cross-country comparison, it may overlook regional and local differences within countries. Educational service demand often varies between urban and rural areas, developed and underdeveloped regions, and different types of institutions. Future studies could apply the proposed framework to provincial, municipal, or school-level datasets to provide more precise evidence for local service planning.

Third, although XGBoost achieves the best predictive performance, ensemble learning models are less transparent than traditional statistical models. Feature importance analysis improves interpretability, but it cannot fully explain causal mechanisms. Future research may incorporate explainable AI methods such as SHAP analysis, partial dependence plots, or counterfactual explanations to better show how specific variables influence predicted educational service demand.

Finally, the current framework is based on historical data, which means that sudden shocks may affect forecasting accuracy. Events such as pandemics, economic crises, demographic changes, education reforms, or technological disruptions may change educational service demand in ways that historical patterns cannot fully capture. Future research may introduce scenario simulation, real-time monitoring data, or dynamic forecasting models to improve adaptability under uncertainty.

Overall, this study demonstrates that AI-based forecasting can provide valuable support for educational service planning when embedded in a broader service informatics framework. The contribution of the study lies not only in showing the superior performance of XGBoost, but also in demonstrating how multi-source public data, ESDI construction, machine learning, and service-system interpretation can be combined to support evidence-based educational decision-making.

6. Conclusions

This study develops an artificial intelligence-based framework for forecasting educational service demand at the country level, integrating multi-source international data, a composite Educational Service Demand Index (ESDI), and machine learning models. By combining linear and nonlinear approaches, including Ridge Regression, Random Forest, and XGBoost, the study demonstrates that ensemble learning can capture complex, nonlinear interactions among enrollment, government expenditure, teacher availability, digital infrastructure, and socioeconomic factors. XGBoost consistently achieves superior predictive performance, highlighting its suitability for analyzing multidimensional educational service systems.

The results provide actionable insights for policymakers and educational administrators. Lagged enrollment rates, public expenditure, and digital readiness are identified as the key drivers of future educational service demand, suggesting that capacity planning, resource allocation, and digital infrastructure investment should be prioritized according to predicted demand. Temporal and regional patterns further indicate the need for differentiated strategies across countries and income levels to ensure equitable access to quality education.

Methodologically, the study illustrates the value of combining composite index construction with machine learning for service-system analysis, offering a reproducible and interpretable forecasting approach. Overall, the proposed framework contributes to both service informatics research and practical educational planning, providing a robust decision-support tool for evidence-based resource allocation, capacity expansion, and digital service improvement across diverse educational contexts.

References

Baker, R. S. J. D. & Yacef, K. (2009), The state of educational data mining in 2009: A review and future visions, Journal of Educational Data Mining, Vol. 1, No. 1, 3-17.

Baker, R. S., Inventado, P. S. (2014), Educational data mining and learning analytics, in Larusson, J. A. & White, B. (Eds.), Learning Analytics: From Research to Practice, Springer, New York, 61-75.

Barrett, M., Davidson, E., Prabhu, J. & Vargo, S. L. (2015), Service innovation in the digital age: Key contributions and future directions, MIS Quarterly, Vol. 39, No. 1, 135-154.

Berland, M., Baker, R. S. & Blikstein, P. (2014), Educational data mining and learning analytics: Applications to constructionist research, Technology, Knowledge and Learning, Vol. 19, No. 1-2, 205-220.

Breiman, L. (2001), Random forests, Machine Learning, Vol. 45, No. 1, 5-32.

Cheng, X. (2026), A service-oriented and informatics-based mobile e-government system for public crisis management, Journal of Logistics, Informatics and Service Science, Vol. 13, No. 6, 17-31.

Chen, T. & Guestrin, C. (2016), XGBoost: A scalable tree boosting system, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, 785-794.

Cornejo-Meza, G., Vivanco, J. & Huayanay, M. (2025), Empowering marketing strategies of SMEs in emerging countries with generative artificial intelligence, Journal of Service, Innovation and Sustainable Development, Vol. 6, No. 2, 28-42.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, New York.

Hoerl, A. E. & Kennard, R. W. (1970), Ridge regression: Biased estimation for nonorthogonal problems, Technometrics, Vol. 12, No. 1, 55-67.

IfM and IBM. (2008), Succeeding through Service Innovation: A Service Perspective for Education, Research, Business and Government, University of Cambridge Institute for Manufacturing, Cambridge.

Jolliffe, I. T. (2002), Principal Component Analysis, 2nd ed., Springer, New York.

Jordan, M. I. & Mitchell, T. M. (2015), Machine learning: Trends, perspectives, and prospects, Science, Vol. 349, No. 6245, 255-260.

Kotsiantis, S. B., Pierrakeas, C. J. & Pintelas, P. E. (2004), Predicting students' performance in distance learning using machine learning techniques, Applied Artificial Intelligence, Vol. 18, No. 5, 411-426.

Long, P. & Siemens, G. (2011), Penetrating the fog: Analytics in learning and education, EDUCAUSE Review, Vol. 46, No. 5, 30-40.

Maglio, P. P. & Spohrer, J. (2008), Fundamentals of service science, Journal of the Academy of Marketing Science, Vol. 36, No. 1, 18-20.

Nambisan, S., Lyytinen, K., Majchrzak, A. & Song, M. (2017), Digital innovation management: Reinventing innovation management research in a digital world, MIS Quarterly, Vol. 41, No. 1, 223-238.

OECD. (2023), Education at a Glance 2023: OECD Indicators, OECD Publishing, Paris.

Ostrom, A. L., Bitner, M. J., Brown, S. W., Burkhard, K. A., Goul, M., Smith-Daniels, V., Demirkan, H. & Rabinovich, E. (2010), Moving forward and making a difference: Research priorities for the science of service, Journal of Service Research, Vol. 13, No. 1, 4-36.

Romero, C. & Ventura, S. (2010), Educational data mining: A review of the state of the art, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 40, No. 6, 601-618.

Spohrer, J., Maglio, P. P., Bailey, J. & Gruhl, D. (2007), Steps toward a science of service systems, Computer, Vol. 40, No. 1, 71-77.

Thotharat, N., Thanawastien, S. & Tantaswadi, P. (2019a), Strategies for the successful implementation of a workflow system for SEM within education, International Journal of Simulation: Systems, Science and Technology, Vol. 20, DOI: 10.5013/IJSSST.a.20.05.16.

Thotharat, N., Thanawastien, S. & Tantaswadi, P. (2019b), Recruitment tracking system and education analytics for strategic enrollment management, International Journal of Simulation: Systems, Science and Technology, Vol. 20, DOI: 10.5013/IJSSST.a.20.05.17.

Thanh, L. D. N., Hai, P. T., Cong, H. T. & Quan, D. D. (2026), Managing hybrid work teams: Leadership strategies for the post-pandemic digital workplace, Journal of Management Changes in the Digital Era, Vol. 3, No. 1, 1-10.

Thapa, N. (2026), A bibliometric analysis of fraud detection in online transactions, Journal of System and Management Sciences, Vol. 16, No. 1, 37-51.

UNESCO. (2015), Education 2030: Incheon Declaration and Framework for Action for the Implementation of Sustainable Development Goal 4, UNESCO, Paris.

UNESCO. (2023), Global Education Monitoring Report 2023: Technology in Education: A Tool on Whose Terms?, UNESCO, Paris.

UNESCO Institute for Statistics. (2024), UIS Data Browser, UNESCO Institute for Statistics, Montreal.

UNICEF and International Telecommunication Union. (2020), How Many Children and Young People Have Internet Access at Home? Estimating Digital Connectivity during the COVID-19 Pandemic, UNICEF, New York.

Vargo, S. L. & Lusch, R. F. (2004), Evolving to a new dominant logic for marketing, Journal of Marketing, Vol. 68, No. 1, 1-17.

World Bank. (2018), World Development Report 2018: Learning to Realize Education's Promise, World Bank, Washington, DC.

World Bank. (2024), Education Statistics: EdStats, World Bank, Washington, DC.

Xue, T., Jhantasana, C. & Seedee, R. (2026), Enterprise informatics capabilities and innovation management effectiveness in Chinese technology-based firms: The roles of innovation process quality and digital transformation, Journal of Logistics, Informatics and Service Science, Vol. 13, No. 6, 184-202.