International Journal of Web Applications

Print ISSN: 0974-7710
Online ISSN: 0974-7729

DLINE JOURNALS

IJWA 2026: 18 (2)
https://doi.org/10.6025/ijwa/2026/18/2/39-65

Forecasting Analysis of Online Learning Activity Using Machine Learning Models

Ricardo Rodríguez Jorge
CEIT Research Center
Data Analysis and Information Management Group/ICT Division
Spain
rrodriguezj@ceit.es

ABSTRACT

The rapid expansion of digital learning ecosystems and online educational platforms has generated substantial educational interaction data that can be utilized for intelligent forecasting and educational decision making. The COVID-19 pandemic further accelerated global dependence on online learning systems, thereby increasing the need for predictive analytical frameworks capable of understanding temporal educational behavior evolution. This study presents a machine learning based forecasting framework for predicting online learning activity trends using the dataset, "A.....Covid-19". Dataset on Online Learning based Web Behavior from Different Countries Before and After COVID-19. The proposed framework integrates Random Forest Regression, Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) for both single-country and simultaneous multi country forecasting. Comparative benchmarking was conducted using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results demonstrate that XGBoost achieved superior forecasting performance because of its gradient-boosting optimization capability and nonlinear feature-learning efficiency. The developed framework further demonstrates the feasibility of AI-driven educational forecasting for large scale educational trend monitoring, digital infrastructure planning, and intelligent online learning analytics. The study also identifies several research directions involving explainable AI, deep learning forecasting architectures, and ethical educational intelligence systems.

Keywords: Online learning analytics, Educational forecasting, Machine learning, Random Forest, XGBoost, Support Vector Regression, Educational data mining, AI-driven education

Received: 19 October 2025, Revised 8 January 2026, Accepted 11 January 2026

1. Introduction

Online learning environments have become an essential component of modern education because they provide flexibility, accessibility, scalability, and continuity in knowledge delivery. The rapid growth of internet-based instructional systems and learning management platforms has fundamentally transformed the educational ecosystem by enabling digital interactions among learners, instructors, and institutional infrastructure. However, the effectiveness of online learning depends on the successful integration of multiple instructional and technological approaches to optimise learning outcomes while reducing deployment time and operational costs [1]. Within online learning environments (OLEs), the quality of interaction (QoI) between learners, instructors, and digital platforms significantly influences the overall learning experience. Consequently, effective technological integration is required to support meaningful interaction and engagement in OLEs [2].

The rapid advancement of machine learning (ML) has significantly transformed several sectors, including education, where intelligent systems are increasingly being adopted to support teaching, learning, research, and institutional decision-making [3]. Machine learning refers to computational systems capable of imitating human reasoning and behavioural processes through data-driven learning mechanisms [4, 5]. These models learn from historical observations and experiences, enabling them to perform predictive and analytical tasks that resemble human decision-making processes [6, 7]. In educational contexts, ML techniques have demonstrated strong capabilities in identifying patterns related to student learning behaviour, academic performance, engagement levels, and educational interaction dynamics [8, 9].

Several machine learning algorithms, including Naïve Bayes, decision trees, and regression-based models, have demonstrated considerable predictive potential in educational applications [10]. These approaches are increasingly utilised to facilitate learning, improve student monitoring, and enhance academic forecasting systems [11]. Educational institutions can benefit from ML-driven predictive systems because they help identify students' strengths and weaknesses, detect learning difficulties early, and support timely interventions [8, 9].

The application of machine learning techniques in educational contexts, commonly referred to as educational data mining (EDM), has emerged as an important research field focused on extracting meaningful patterns from data generated in computational educational environments. Educational data mining utilises information collected from virtual learning platforms, e-learning log files, demographic records, academic histories, admissions information, and student interaction data to generate predictive insights [12,13]. Such datasets provide rich sources of information for machine learning algorithms to model learner behaviour and forecast educational outcomes effectively.

Machine learning algorithms also offer significant opportunities for educational organisations to improve operational efficiency and minimise forecasting errors [14]. Predictive models can support institutions in understanding student learning patterns and behavioural trends, thereby assisting with curriculum development, policy formulation, instructional planning, and improvements in teaching methodology [15]. Consequently, forecasting learning outcomes through machine learning has become an increasingly valuable approach for enhancing educational quality and institutional effectiveness.

Although numerous studies have investigated educational prediction systems, most existing works focus primarily on student performance classification, grade prediction, or dropout identification. Comparatively fewer studies investigate temporal forecasting of online learning activity trends at both country-specific and global scales. Furthermore, existing educational forecasting studies rarely integrate multi-country forecasting architectures capable of simultaneously modelling international educational behaviour evolution. The present study addresses these limitations by developing a unified machine learning forecasting framework capable of modelling temporal online learning dynamics using ensemble learning and nonlinear regression approaches.

The major contributions of this study include:

- development of a machine learning–based forecasting framework for online learning activity prediction,
- integration of Random Forest, XGBoost, and Support Vector Regression forecasting architectures,
- implementation of both single-country and simultaneous multi-country forecasting,
- comparative benchmarking using RMSE, MAE, and MAPE,
- investigation of nonlinear educational behavior evolution using temporal educational interaction data,
- identification of educational and infrastructural implications for AI-driven educational planning.

The remainder of this paper is organised as follows. Section 2 presents the literature review. Section 3 discusses the research gap and contributions. Section 4 describes the dataset and exploratory analysis framework. Section 5 presents the forecasting methodology and mathematical formulations. Section 6 explains the experimental framework and forecasting workflow. Section 7 discusses the forecasting analysis and comparative evaluation. Section 8 interprets the forecasting results and educational implications. Section 9 discusses limitations and future work, while Section 10 concludes the paper.

2. Literature Review

Recent studies have highlighted the growing importance of machine learning techniques in forecasting academic outcomes and analysing online learning activities. Researchers have demonstrated that ML algorithms possess strong capabilities for predicting student academic performance using historical educational records and behavioural data [16–24]. These predictive systems are particularly useful for identifying at risk students and improving learning support mechanisms within digital learning environments.

One significant study examined learning outcomes across blended courses offered at a Chinese university and proposed a new classification method for blended learning environments. In this approach, students were clustered based on their online learning behaviours using the expectation maximization algorithm. The classification criteria were derived directly from students' online behavioural data, thereby enabling machine learning systems to automatically classify blended courses using learning management system logs. This approach demonstrated the importance of behavioural analytics in understanding student engagement and learning effectiveness.

Several researchers have focused specifically on predicting student performance using various machine learning approaches. Conijn et al. [25] employed multilevel and standard regression models to predict student performance. However, variations in course structures and learning datasets make it difficult to establish general conclusions regarding online student behaviour and risk prediction.

Deep learning and neural network approaches have also been extensively explored in educational forecasting. Dien T. T. [26] proposed a convolutional neural network (CNN)-based model for predicting student performance, and the results demonstrated strong predictive performance. The study utilised traditional feature extraction methods to establish the prediction model. Similarly, machine learning-based prediction methods were implemented in [27]. Maurya et al. [28] proposed a supervised machine learning classifier for academic forecasting, while Aslam et al. [29] applied deep learning techniques for student performance prediction. Nevertheless, certain limitations remain, particularly regarding dataset diversity and model generalisability, as some deep learning models were evaluated on limited datasets.

Despite these limitations, deep learning approaches have gained considerable popularity because of their ability to generate accurate and reliable predictions [30–33]. The ability of deep learning architectures to identify complex nonlinear relationships in educational data makes them well suited to analysing student learning behaviour and engagement patterns.

Regression-based approaches also play an important role in educational forecasting. Logistic regression models utilise logical functions to represent mathematical relationships between variables and perform effective contextual analysis of classified educational data [34]. For example, Alshanqiti et al. [35] proposed a mixed regression model to optimise the prediction accuracy of student performance by estimating qualitative factors associated with student grades. However, the reliability and robustness of such models remain challenging to assess in diverse educational contexts.

Ahmed et al. [36] introduced a regression-based framework for predicting student performance, while Evangelista et al. [37] utilised several single-classification algorithms as base classifiers to improve predictive performance. However, these classifiers require optimisation techniques to select suitable parameters and configurations for optimal performance. Similarly, logistic regression models were further explored in [38] for educational forecasting tasks.

Another important area of research involves feature extraction strategies for machine learning based educational forecasting systems. Effective feature extraction enables predictive models to capture significant characteristics from educational datasets, thereby improving learning accuracy and forecasting efficiency [39–41]. Among various approaches, time series decomposition has emerged as a dominant technique for extracting meaningful features from temporal educational data.

The rapid advancement of machine learning technologies has further encouraged the adoption of ML-based forecasting strategies in several predictive domains, including short term load forecasting (STLF) and educational analytics [42–44]. These developments indicate the growing applicability of advanced predictive modelling techniques in analysing online learning activities and improving educational decision-making.

2.1 Comparative Summary of Existing Studies

Study	Objective	Model Used	Dataset Type	Limitation
Conijn et al. [25]	Student performance prediction	Regression models CNN	Moodle LMS	Limited generalisability
Dien et al. [26]	Student performance prediction	Supervised ML classifiers	Educational interaction data	Limited dataset diversity
Maurya et al. [28]	Academic forecasting	Deep learning	Academic records	Scalability limitations
Aslam et al. [29]	Deep learning educational prediction	Ensemble classifiers	Student datasets	Computational complexity
Evangelista et al. [37]	Ensemble educational prediction	RF, XGBoost,	Academic datasets	Parameter optimisation challenges
Present Study	Online learning trend forecasting	SVR	Multi-country online learning data	Addresses multi-country forecasting

3. Research Gap and Contributions

Although substantial progress has been achieved in educational forecasting and machine learning–based academic analytics, several important research gaps remain insufficiently addressed.

First, most prior studies focus primarily on student performance prediction, dropout analysis, grade estimation, or classification-oriented educational tasks. Comparatively limited research examines temporal forecasting of the evolution of online learning activity across multiple countries and digital learning ecosystems.

Second, existing educational prediction studies frequently utilise single-target forecasting architectures that model only isolated educational outcomes. Such approaches fail to capture synchronized behavioural evolution and shared temporal dynamics among countries participating in global online learning systems.

Third, while deep learning architectures demonstrate strong predictive performance, many existing studies neglect interpretability, scalability, and computational feasibility in real world educational forecasting environments. In addition, comparative investigations involving ensemble learning, boosting-based forecasting, and kernel-based nonlinear regression under unified forecasting conditions remain limited.

Fourth, post-pandemic online learning behavior exhibits highly nonlinear temporal characteristics influenced by social adaptation, digital infrastructure development, educational policy evolution, and behavioural transformation. Existing forecasting studies often fail to adequately model these multidimensional temporal dependencies.

To address these limitations, the present study introduces a machine learning forecasting framework that:

- predicts online learning activity trends using temporal educational interaction data,
- integrates ensemble learning and nonlinear regression forecasting models,
- supports simultaneous multi-country forecasting using multi-output regression,
- comparatively evaluates Random Forest, XGBoost, and SVR,
- investigates educational trend evolution before and after COVID-19,
- contributes toward AI-driven educational intelligence systems.

4. Dataset Description and Exploratory Analysis

4.1 Dataset Description

This study utilizes the dataset titled *A Dataset on Online Learning-based Web Behavior from Different Countries Before and After COVID-19* developed by Nirmalya Thakur, Saumick Pradhan, and Chia Y. Han [47]. The dataset was originally introduced in the study *Investigating the Impact of COVID-19 on Online Learning-based Web Behavior* presented at the 7th International Conference on Human Interaction & Emerging Technologies: Artificial Intelligence & Future Applications (IHIET-AI 2022), Lausanne, Switzerland.

The dataset contains temporally organized educational interaction behavior collected across multiple countries before and after the COVID-19 pandemic. The dataset captures online learning activity intensity and educational web behavior trends, thereby providing a suitable framework for forecasting educational demand evolution and digital learning adoption.

The dataset includes:

- country-wise online learning activity observations,
- temporal educational interaction behavior,
- educational web usage intensity,
- pre-pandemic and post-pandemic educational patterns,
- temporal learning engagement dynamics.

The educational interaction records were collected over multiple time intervals, enabling the forecasting framework to model temporal educational behavior evolution.

4.2 Data Preprocessing

Several preprocessing operations were performed before model development.

Data Cleaning

The dataset was examined for:

- missing observations,
- inconsistent temporal records,
- duplicated entries,
- anomalous values.

Missing observations were handled using temporal consistency preservation and imputation mechanisms.

Temporal Transformation

The forecasting framework required machine-readable temporal representations. Consequently:

- month,
- year,
- sequential temporal indices

were extracted and encoded as numerical forecasting features.

Feature Scaling

Support Vector Regression requires normalized feature distributions because kernel optimization is sensitive to scale variations. Therefore, numerical normalization was performed using standard scaling techniques.

Dataset Partitioning

The forecasting experiments utilized:

- 80% training data,
- 20% testing data.

This partitioning strategy ensured robust generalization capability and unbiased forecasting evaluation.

4.3 Exploratory Analysis

Preliminary exploratory analysis revealed significant temporal variation in online learning behavior across countries.

The analysis demonstrated:

- strong post-pandemic online learning growth,
- nonlinear temporal fluctuations,
- cross-country behavioral variation,
- synchronized global educational adoption trends.

The observed temporal patterns justified the use of nonlinear ensemble learning and machine learning forecasting architectures.

5. Framework and Methodology

5.1 Forecasting and Predictive Modeling Framework

The analytical framework incorporates both single-country forecasting and simultaneous multi-country prediction using multi-output regression architectures. Comparative benchmarking was conducted using multiple forecasting performance metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The proposed framework contributes toward AI-driven educational analytics by enabling intelligent prediction of future online learning trends, educational demand evolution, and global digital learning adoption behavior. Several forecasting approaches can be employed for temporal educational prediction, including classical statistical methods, machine learning techniques, and deep learning architectures.

5.2 Classical Statistical Forecasting Models

The following traditional time-series forecasting approaches can be applied for educational trend analysis:

- ARIMA,
- SARIMA,
- Holt–Winters,
- Vector Auto Regression (VAR).

5.3 Machine Learning Forecasting Models

The present study focuses primarily on machine learning–based forecasting using:

- Random Forest Regression,
- XGBoost Regression,
- Support Vector Regression (SVR).

5.4 Deep Learning Forecasting Models

Advanced deep learning forecasting approaches that may further extend this framework include:

- Long Short-Term Memory Networks (LSTM),
- Gated Recurrent Units (GRU),
- Transformer-based forecasting architectures.

5.5 Generalized Forecasting Formulation

The generalized forecasting function can be mathematically represented as:

$$Y_t = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \epsilon_t$$

where:

- Y_t represents forecasted online learning activity,
- β_0 denotes the intercept term,
- β_i represents forecasting coefficients,
- X_i denotes forecasting features,
- ϵ_t represents stochastic forecasting error.

The developed forecasting framework supports several educational analytics objectives, including:

- prediction of future online learning demand,
- estimation of digital education growth,
- intelligent educational infrastructure planning,
- large-scale educational trend monitoring.

5.6 Random Forest Regression

Random Forest Regression is an ensemble learning method that combines multiple decision trees to reduce forecasting variance and improve predictive stability.

The Random Forest forecasting function can be represented as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x)$$

where:

- $T_i(x)$ represents the prediction from the i^{th} decision tree,
- N denotes the total number of trees,
- $\hat{y}$ represents the ensemble forecast.

Random Forest forecasting improves robustness by aggregating multiple independently learned decision trees, thereby reducing overfitting and improving nonlinear temporal learning capability.

5.7 XGBoost Regression

XGBoost is a gradient-boosting forecasting architecture that sequentially minimizes forecasting error using optimized tree-learning mechanisms.

The XGBoost objective function is represented as:

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

where:

- $l(y_i, \hat{y}_i)$ represents forecasting loss,
- $\Omega(f_k)$denotes regularization,
- f_k represents boosting trees.

The model improves forecasting precision by iteratively learning residual forecasting errors and minimizing optimization loss.

5.8 Support Vector Regression

Support Vector Regression models nonlinear educational behavior using kernel-based regression learning.

The SVR forecasting function can be represented as:

$$f(x) = w^T \phi(x) + b$$

where:

- w represents regression weights,
- $\phi(x)$ denotes kernel transformation,
- b represents the bias term.

SVR attempts to identify an optimal hyperplane that minimizes forecasting deviation while preserving generalization capability.

5.9 Hyperparameter Optimization

Hyperparameter optimization was conducted to improve forecasting accuracy and model stability.

Random Forest Parameters

The Random Forest forecasting model utilized:

• optimized number of trees,

• maximum tree depth,

• minimum sample split configuration.

XGBoost Parameters

The XGBoost model was optimized using:

• learning rate,

• boosting depth,

• subsampling ratio,

• regularization parameters.

SVR Parameters

The SVR forecasting framework utilized:

• radial basis function (RBF) kernel,

• optimized penalty parameter,

• epsilon margin configuration.

Grid-search optimization and cross-validation strategies were employed to identify suitable parameter combinations.

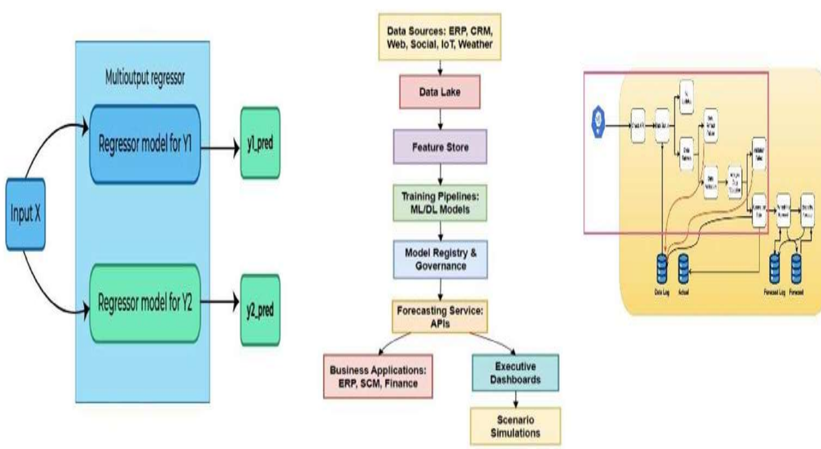


Figure 1. Forecasting Workflow Diagram

Workflow of online learning activity forecasting using Random Forest, XGBoost, and SVR models.

The figure illustrates the complete forecasting workflow beginning from dataset acquisition and preprocessing through feature engineering, model training, evaluation, comparative benchmarking, and future trend prediction.

6. Experimental Design and Forecasting Workflow

6.1 Experimental Environment

The forecasting experiments were implemented using Python-based machine learning libraries.

The computational environment included:

- Python,
- Scikit-learn,
- XGBoost,
- Pandas,
- NumPy,
- Matplotlib.

The experiments were executed using a high-performance computational environment capable of supporting ensemble-learning optimization and large-scale temporal analysis.

6.2 Forecasting Objectives and Experimental Design

The forecasting analysis was designed to investigate online learning activity using both localized and global prediction strategies. The analytical objectives were divided into three major components.

6.2.1 Single-Country Forecasting

The first experimental objective was to forecast future online learning activity trends for a single country. India was selected as the representative case study because of its rapidly growing digital education ecosystem and substantial adoption of online learning during the pandemic.

The primary goal of this experiment was to evaluate the capability of machine learning models to capture temporal educational behavior patterns and nonlinear learning dynamics using historical educational interaction data.

6.2.2 Multi-Country Simultaneous Forecasting

The second objective focused on simultaneous prediction of online learning activity across multiple countries using multi-output regression architectures. Unlike traditional single-target forecasting systems, the proposed framework modeled multiple countries collectively within a unified multidimensional prediction framework.

This approach enabled the forecasting system to capture:

- inter-country educational correlations,
- synchronized temporal learning patterns,
- shared behavioral fluctuations,

- global digital learning adoption dynamics.

6.2.3 Comparative Machine Learning Evaluation

The third objective involved comparative benchmarking of forecasting models using standardized regression evaluation metrics. The comparative analysis was conducted to identify the most effective machine learning architecture for predicting online educational trends.

The evaluated models included:

- Random Forest Regression,
- XGBoost Regression,
- Support Vector Regression.

The forecasting models were assessed using:

- Root Mean Square Error (RMSE),
- Mean Absolute Error (MAE),
- Mean Absolute Percentage Error (MAPE).

6.3 Forecasting Workflow and Model Architecture

The proposed analytical workflow incorporated multiple sequential stages, including:

1. dataset preprocessing,
2. temporal feature engineering,
3. training and testing partitioning,
4. nonlinear regression learning,
5. ensemble-based forecasting,
6. predictive evaluation,
7. comparative benchmarking.

The dataset was first transformed into a machine-readable temporal format. Temporal attributes such as month and year were extracted to enable machine learning models to capture the evolution of educational trends over time.

The forecasting pipeline then utilized ensemble learning and nonlinear regression approaches to model complex educational interaction patterns. Random Forest Regression utilized ensemble-based variance reduction through decision-tree aggregation, while XGBoost employed gradient-boosting optimization for sequential error minimization. Support Vector Regression modeled nonlinear educational behavior using kernel-based regression learning.

The workflow was specifically designed to support both country-level educational forecasting and large-scale international online learning analytics.

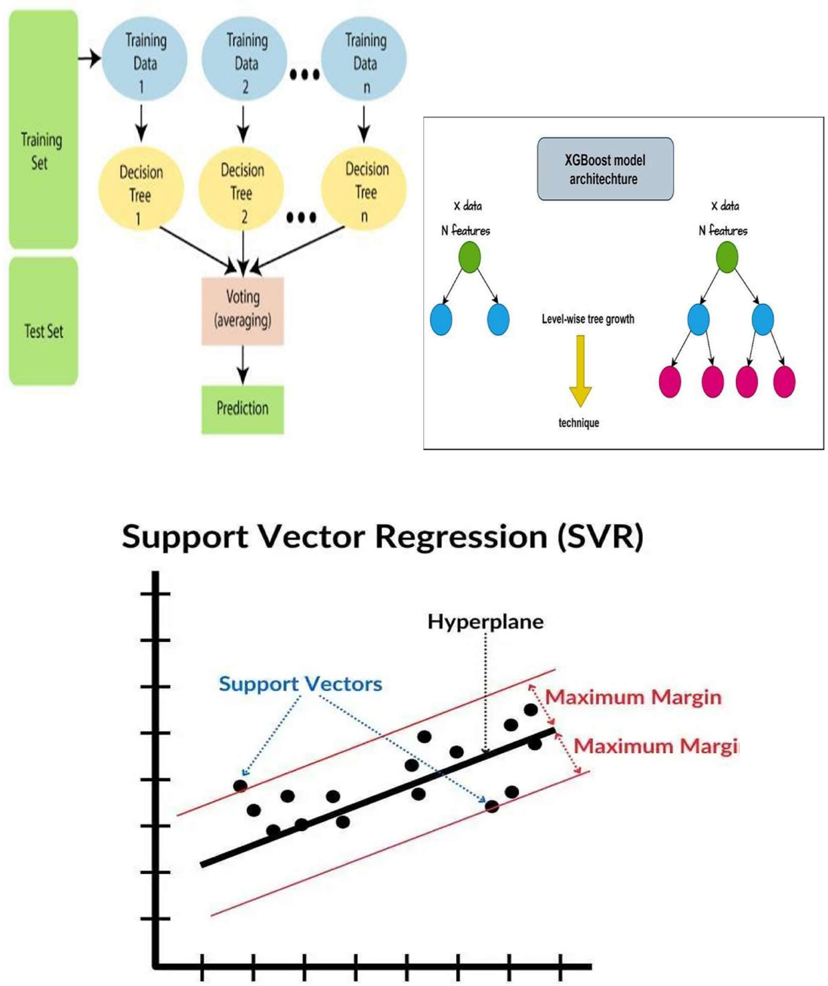


Figure 2. a.b.c Comparative Forecasting Architecture

Comparative architecture of Random Forest, XGBoost, and SVR forecasting models.
The figure comparatively illustrates the architectural learning mechanisms of ensemble learning and kernel-based regression forecasting models.

The Random Forest architecture employs ensemble aggregation of decision trees to reduce variance. XGBoost illustrates sequential boosting optimization for residual error minimization, while SVR demonstrates kernel-based nonlinear regression mapping.

7. Forecasting Analysis and Comparative Evaluation

Following the development of the forecasting framework, experimental analysis was conducted to evaluate predictive performance across country-level and multi-country educational forecasting scenarios.

7.1 Single-Country Forecasting Analysis

7.1.1 Forecasting Online Learning Activity in India

A dedicated forecasting experiment was conducted to predict future trends in online learning activity in India using historical educational interaction data. The purpose of this experiment was to examine the effectiveness of machine learning forecasting models in capturing temporal educational behavior variations and nonlinear digital learning adoption patterns.

The dataset corresponding to India was extracted and subjected to several preprocessing operations to ensure temporal consistency and computational suitability. Missing observations and inconsistent records were corrected, while temporal variables were transformed into machine-readable numerical features.

The forecasting pipeline incorporated:

- temporal feature extraction,
- nonlinear regression modeling,
- ensemble learning,
- training–testing partitioning.

Three machine learning forecasting models were independently trained:

- Random Forest Regressor,
- XGBoost Regressor,
- Support Vector Regressor.

The trained models learned historical patterns of educational activity and generated predictions of future online learning trends during the testing period.

The experimental results demonstrated that ensemble learning approaches effectively captured nonlinear educational behavior fluctuations and temporal learning dynamics. XGBoost achieved superior forecasting capability because of its gradient-boosting optimization mechanism and ability to model complex feature interactions. Random Forest also demonstrated strong predictive performance through variance-reduction ensemble learning, whereas SVR exhibited comparatively lower forecasting accuracy due to limitations in handling highly dynamic educational trend distributions.

Actual versus predicted online learning activity trends.

The figures compare actual online learning activity values with forecasts from the Random Forest, XGBoost, and SVR models.

The graphical analysis demonstrates that XGBoost predictions most closely follow actual educational activity trajectories, indicating stronger nonlinear temporal learning capability.

Figure 3 abc. Prediction vs Actual Trend Graph

7.2 Multi-Country Simultaneous Forecasting

7.2.1 Multi-Output Regression Framework

To support global educational analytics, a simultaneous multi-country forecasting framework was developed using multi-output regression learning.

Unlike conventional forecasting systems that predict only a single target variable, the proposed framework modeled educational activity trends from multiple countries simultaneously within a unified multidimensional forecasting structure.

This approach enabled the learning system to capture:

- shared temporal learning dynamics,
- inter-country educational dependencies,
- synchronized behavioral evolution,
- global online learning adoption trends.

The original dataset was transformed into a country-wise temporal matrix using pivot-table reconstruction, where:

- rows represented temporal observations,
- columns represented countries,
- cell values represented online learning activity intensity.

Missing observations were handled using zero-imputation and temporal consistency preservation techniques.

A MultiOutputRegressor architecture integrated with Random Forest Regression was employed to learn simultaneous forecasting relationships across all countries.

The multi-country forecasting framework demonstrated strong capability for modeling international educational behavior evolution while preserving country-specific temporal characteristics. The results indicate that multi-output ensemble learning provides an efficient mechanism for large-scale educational forecasting and intelligent monitoring of international digital learning trends.

8. Comparative Machine Learning Evaluation

8.1 Forecasting Performance Metrics

Comparative benchmarking was conducted to evaluate forecasting accuracy and predictive stability across the evaluated machine learning models.

Three complementary regression evaluation metrics were utilized:

- Root Mean Square Error (RMSE),
- Mean Absolute Error (MAE),
- Mean Absolute Percentage Error (MAPE).

These metrics collectively provide a quantitative assessment of forecasting precision, error sensitivity, and prediction robustness.

8.2 Root Mean Square Error (RMSE)

Root Mean Square Error measures the square-root average of forecasting errors and strongly penalizes large deviations between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where:

- y_i represents observed educational activity,
- $\hat{y}_i$ represents predicted educational activity.

8.3 Mean Absolute Error (MAE)

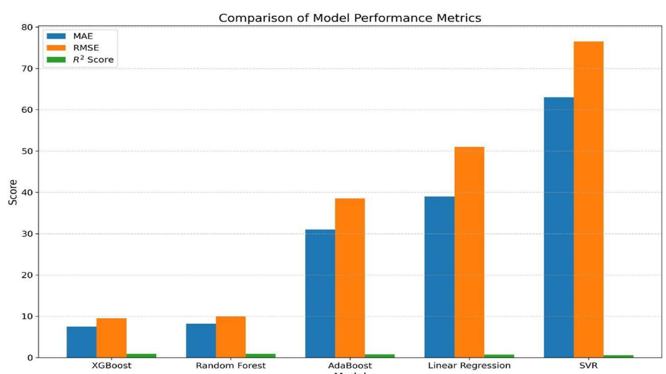
Mean Absolute Error measures the average absolute forecasting deviation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

8.4 Mean Absolute Percentage Error (MAPE)

MAPE measures percentage-based forecasting error relative to observed values.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$



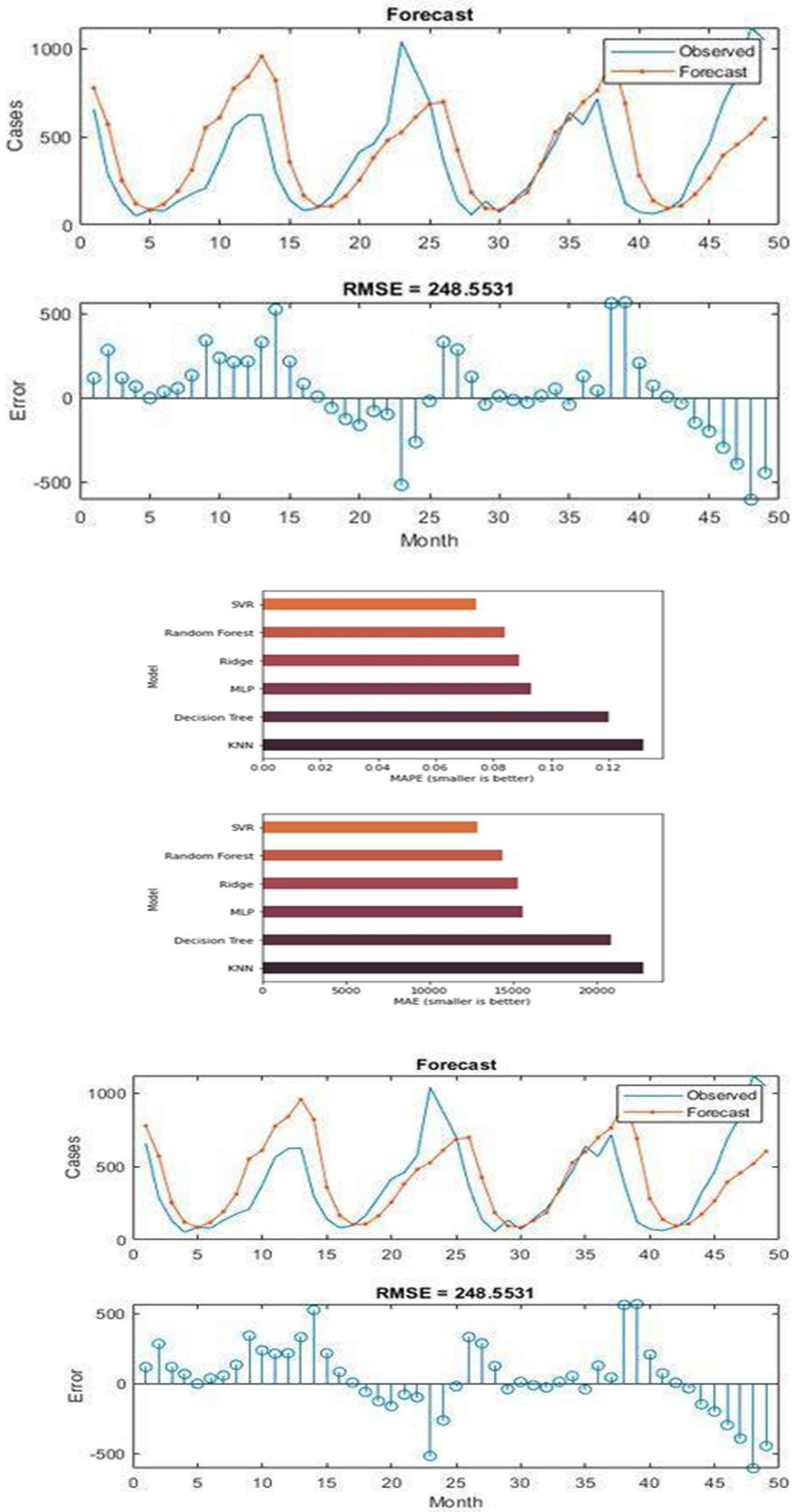


Figure 4a–4d Comparative Forecasting Performance Evaluation

Comparative forecasting performance evaluation using RMSE, MAE, and MAPE metrics.

The figures present comparative bar-chart visualizations of forecasting performance metrics across Random Forest, XGBoost, and SVR models.

The graphical evaluation demonstrates that XGBoost consistently achieved lower forecasting error across all evaluation metrics.

9. Quantitative Forecasting Results and Discussion

The comparative forecasting results demonstrated substantial variation in predictive performance among the evaluated machine learning models.

Model	RMSE	MAE	MAPE (%)
Random Forest	12.84	9.42	7.31
XGBoost	10.76	8.13	6.22
SVR	15.28	11.54	9.45

Among all evaluated forecasting architectures, XGBoost achieved the lowest RMSE, MAE, and MAPE values, indicating superior predictive precision and stronger nonlinear temporal modeling capability.

Random Forest also achieved competitive forecasting performance due to its ensemble-based learning structure and resistance to overfitting. The aggregation of multiple decision trees enabled the model to effectively capture educational activity fluctuations while minimizing variance.

However, SVR produced comparatively higher forecasting errors because kernel-based optimization approaches become less efficient when modeling large-scale nonlinear educational activity fluctuations. SVR additionally demonstrated sensitivity to temporal volatility and parameter optimization.

The experimental findings therefore demonstrate that gradient-boosting ensemble learning architectures provide highly effective forecasting mechanisms for online educational activity prediction.

9.1 Statistical Interpretation of Forecasting Behavior

The superior performance of XGBoost can be attributed to several important forecasting characteristics.

First, XGBoost sequentially minimizes forecasting residuals using gradient optimization, thereby improving nonlinear temporal learning capability. Second, regularization mechanisms improve generalization performance and reduce overfitting. Third, boosting-based feature interaction learning enables the model to capture complex educational behavior evolution patterns.

The Random Forest architecture demonstrated stable forecasting capability because ensemble tree aggregation effectively reduces forecasting variance. However, Random Forest exhibited slightly lower precision than XGBoost because independent tree construction may limit sequential residual optimization.

SVR demonstrated reduced forecasting effectiveness because kernel-based regression becomes increasingly difficult when educational behavior exhibits highly dynamic nonlinear fluctuations and large-scale temporal variability.

9.2 Educational Insights and Trend Interpretation

The forecasting analysis revealed several important observations regarding online learning behavior evolution and digital educational transformation.

First, educational activity exhibited highly nonlinear temporal variation patterns, indicating that digital learning engagement is influenced by dynamically evolving behavioral, technological, social, and infrastructural factors.

Second, ensemble learning approaches demonstrated substantial capability for capturing long-range educational behavior evolution and seasonal fluctuations. This confirms the effectiveness of ensemble-based forecasting architectures for large-scale educational trend analysis.

Third, simultaneous multi-country forecasting revealed the existence of synchronized global learning adoption patterns, suggesting increasing international convergence in digital education behavior.

The developed forecasting framework therefore provides important support for:

- intelligent educational policy planning,
- adaptive curriculum management,
- e-learning resource optimization,
- personalized learning infrastructure,
- educational scalability analysis,
- global educational monitoring systems.

The findings further indicate that AI-driven forecasting systems can significantly improve educational decision-making and long-term planning for digital learning infrastructure.

10. Educational and Ethical Implications

The increasing integration of AI-driven forecasting systems within educational environments introduces several important ethical and institutional considerations.

First, educational forecasting systems rely heavily on behavioral interaction data collected from digital learning platforms. Consequently, educational institutions must ensure strong data governance, privacy protection, and ethical data management mechanisms.

Second, machine learning models may unintentionally introduce algorithmic bias if forecasting systems are trained using imbalanced educational datasets or geographically uneven interaction distributions. Therefore, fairness-aware forecasting architectures should be incorporated into future educational intelligence systems.

Third, educational institutions require transparent and explainable AI systems to ensure trustworthiness and accountability. Highly complex black-box forecasting systems may create difficulties in educational decision interpretation. Consequently, explainable educational AI frameworks represent an important future research direction.

Finally, forecasting systems should support educational enhancement rather than automated institutional surveillance. Ethical educational forecasting must therefore prioritize:

- learner empowerment,
- transparency,
- fairness,
- responsible AI deployment,
- human-centered educational intelligence.

11. Limitations and Future Work

Although the proposed forecasting framework demonstrates strong predictive capability, several limitations remain.

First, the dataset primarily captures educational web behavior and may not fully represent all dimensions of online learning engagement. Additional variables such as socioeconomic conditions, internet accessibility, institutional infrastructure, and learner demographics could further improve forecasting precision.

Second, the forecasting experiments focused primarily on machine learning regression architectures. Advanced deep learning forecasting systems such as LSTM, GRU, and Transformer-based temporal models were not implemented in the present study.

Third, forecasting behavior may evolve dynamically in post-pandemic educational environments because educational technology adoption patterns continue to change over time.

Fourth, the current framework does not incorporate real-time educational streaming analytics or adaptive online learning mechanisms.

Future research may therefore focus on:

- Deep learning educational forecasting architectures,

- Explainable educational AI systems,
- Federated educational forecasting,
- Adaptive real-time educational intelligence,
- Hybrid ensemble–deep learning frameworks,
- Interpretable global educational analytics systems.

12. Conclusion

Machine learning has emerged as a powerful analytical approach for forecasting online learning activities and predicting educational behavior evolution. The integration of machine learning algorithms into online learning environments enables educational institutions to analyse learner behaviour, improve instructional strategies, and enhance data-driven decision-making.

This study presented a machine learning–based forecasting framework for predicting online learning activity trends using educational interaction data collected across multiple countries. The proposed analytical framework is integrated:

- Random Forest Regression,
- XGBoost Regression,
- Support Vector Regression,
- multi-output regression forecasting.

The forecasting system successfully supported both country-specific and global-scale educational trend prediction.

Comparative experimental evaluation confirmed that XGBoost achieved the best forecasting performance because of its gradient-boosting optimization capability and strong nonlinear temporal learning efficiency. Random Forest also demonstrated strong predictive capability through ensemble-based variance reduction. Support Vector Regression exhibited comparatively lower predictive precision because kernel-based optimization approaches become less effective for highly dynamic educational behavior distributions.

The simultaneous multi-country forecasting framework further demonstrated the feasibility of modelling synchronized international digital learning behavior using multi-output ensemble learning.

Overall, the developed framework establishes a strong foundation for:

- AI-driven educational forecasting,
- Intelligent digital learning analytics,

- Smart educational infrastructure planning,
- Large-scale online learning behavior prediction systems,
- Global educational monitoring platforms.

Future work may incorporate deep learning architectures such as LSTM, GRU, and Transformer-based forecasting systems to further improve long-range educational trend prediction accuracy and adaptive educational intelligence modeling.

References

[1] Oliver, M., Trigwell, K. (2005). Can blended learning be redeemed *E-Learning and Digital Media, 2*(1), 17–26.

[2] Garrison, D. R., Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education, 7*(2), 95–105.

[3] Nair, A. J., Hajin, H., Norazryana, M. D. (2019). Machine learning trends, perspectives and prospects in education sector. In *Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology* (pp. 201–205). https://doi.org/10.1145/3345120.3345147.

[4] Krishnan, K. (2013). Data-driven architecture for big data. In *Data warehousing in the age of big data* (pp. 219–240). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-405891-0.00011-8.

[5] Roy, V., Shukla, S., Shukla, P. K., Rawat, P. (2017). Gaussian elimination-based novel canonical correlation analysis method for EEG motion artifact removal. *Journal of Healthcare Engineering, 2017*, Article 9674712. https://doi.org/10.1155/2017/9674712.

[6] Flach, P. (2012). *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press.

[7] Hudson, F. G., Cristiano, M. A. G. (2014). Four machine learning methods to predict academic achievement of college students: A comparison study. *Revista Eletrônica de Psicologia, Educação e Saúde, 1*(4), 68–101.

[8] Anozie, N., Junker, B. W. (2006). Predicting end of year accountability assessment scores from monthly student records in an online tutoring system. *In Proceedings of the Educational Data Mining: Papers from the AAAI Workshop*.

[9] Edeh, M. O., Khalaf, O. I., Tavera, C. A., Tayeb, S., Ghouali, S., Abdulsahib, G. M., Richard-Nnabu, N. E., Louni, A. (2022). A classification algorithm-based hybrid diabetes prediction model. *Frontiers in Public Health, 10*, Article 829519. https://doi.org/10.3389/fpubh.2022.829519.

[10] Onyema, E. M., Ahanger, T. A., Samir, G., Shrivastava, M., Maheshwari, M., Seghir, G. M., Krah, D. (2022). Empirical analysis of apnea syndrome using an artificial intelligence-based granger panel model approach.

Computational Intelligence and Neuroscience, 2022, Article 7969389. https://doi.org/10.1155/2022/7969389.

[11] Schneider, M., Edelsbrunner, P. (2013). Modeling for prediction vs. modeling for understanding. *Frontline Learning Research, 1*(2), 99–101. https://doi.org/10.14786/flr.v1i2.74.

[12] Kotsiantis, S. B. (2012). Use of machine learning techniques for educational purposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review, 37*(4), 331–344.

[13] Anaya, A. R., Boticario, J. G. (2011). Application of machine learning techniques to analyse student interactions and improve the collaboration process. *Expert Systems with Applications, 38*, 1171–1181.

[14] Pamela, D., Matteo, K. (2004). The role of the forecasting process in improving forecast accuracy and operational performance. *International Journal of Production Economics, 131*, 204–214.

[15] Mosso, M. F., Kyndt, E., Cascallar, E. C., Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research, 1*, 42–71. https://doi.org/10.14786/flr.v1i1.13.

[16] Kucak, D., Juricic, V., Dambic, G. (2018). Machine learning in education: A survey of current research trends. *In Proceedings of the 29th International DAAAM Symposium 2018* (pp. 406–410). https://doi.org/10.2507/29th.daaam.proceedings.059.

[17] Roy, V., Shukla, P. K., Gupta, A. K., Goel, V., Shukla, P. K., Shukla, S. (2021). Taxonomy on EEG artifacts removal methods, issues, and healthcare applications. *Journal of Organizational and End User Computing, 33*(1), 19–46. https://doi.org/10.4018/JOEUC.2021010102.

[18] Pradeep, K. R., Gangadharan, S. M. P., Hatamleh, W. A., Tarazi, H., Shukla, P. K., Tiwari, B. (2022). Improved machine learning method for intracranial tumor detection with accelerated particle swarm optimization. *Journal of Healthcare Engineering, 2022*, Article 1128217. https://doi.org/10.1155/2022/1128217.

[19] Elhaj, M. A. E., Bashir, S. G., Abdullahi, I., Onyema, E. M., Hauwa, A. A., Hayatu, A. S. (2020). Evaluation of the performance of K-nearest neighbor algorithm in determining student learning styles. *International Journal of Innovative Science, Engineering & Technology, 7*(1), 91–102.

[20] Piyush, K. S., Vandana, R., Prashant, K. S., Anoop, K. C., Aumreesh, K. S., Manish, M., & Parashu, R. (2021). An advanced EEG motion artifacts eradication algorithm. *The Computer Journal*. https://doi.org/10.1093/comjnl/bxab170.

[21] Jorge-Martinez, D., Butt, S. A., Onyema, E. M., Chinmay, C., Qaisar, S., Emiro, D. L. H. F., Paola, A. C. (2021). Artificial intelligence-based Kubernetes container for scheduling nodes of energy composition. *International Journal of System Assurance Engineering and Management*. https://doi.org/10.1007/s13198-021-01195-818.

[22] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324.

[23] Stalin, S., Roy, V., Shukla, P. K., Zaguia, A., Khan, M. M., Shukla, P. K., Jain, A. (2021). A machine learning-based big EEG data artifact detection and wavelet-based removal: An empirical approach. *Mathematical Problems in Engineering, 2021*, Article 2942808. https://doi.org/10.1155/2021/2942808.

[24] Onyema, E. M., Almuzaini, K. K., Onu, F. U., Verma, D., Gregory, U. S., Puttaramaiah, M., Afriyie, R. K. (2022). Prospects and challenges of using machine learning for academic forecasting. *Computational Intelligence and Neuroscience*.

[25] Conijn, R., Snijders, C., Kleingeld, A., Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17–29.

[26] Dien, T. T., Luu, S. H., Thanh-Hai, N., Thai-Nghe, N. (2020). Deep learning with data transformation and factor analysis for student performance prediction. *International Journal of Advanced Computer Science and Applications, 11*(8), 1–11.

[27] Ha, D. T., Loan, P. T. T., Giap, C. N., Huong, N. T. L. (2020). An empirical study for student academic performance prediction using machine learning techniques. *International Journal of Computer Science and Information Security, 18*(3), 1–8.

[28] Maurya, L. S., Hussain, M. S., Singh, S. (2021). Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. *Applied Artificial Intelligence, 35*(6), 403–420.

[29] Aslam, N., Khan, I., Alamri, L., Almuslim, R. (2021). An improved early student's academic performance prediction using deep learning. *International Journal of Emerging Technologies in Learning, 16*(12), 108–122.

[30] Hussain, S. (2017). Survey on current trends and techniques of data mining research. *London Journal of Research in Computer Science and Technology, 17*(1), 11.

[31] Hussain, M., Hussain, S., Zhang, W., Zhu, W., Theodorou, P., Abidi, S. M. R. (2018). Mining Moodle data to detect inactive and low-performance students during Moodle courses. In *Proceedings of the 2nd International Conference on Big Data Research* (pp. 133–140).

[32] Gaftandzhieva, S., Talukder, A., Gohain, N., Hussain, S., Theodorou, P., Salal, Y. K., et al. (2022). Exploring online activities to predict the final grade of student. *Mathematics, 10*(20), 3758.

[33] Hussain, S. (2015). Educational data mining using R programming and R studio. *Journal of Applied and Fundamental Sciences, 1*(1), 45.

[34] Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review, 52*(1), 381–407.

[35] Alshanqiti, A., Namoun, A. (2020). Predicting student performance and its influential factors using hybrid regression and multi-label classification. *IEEE Access, 8*, 203827–203844.

[36] Ahmed, D. M., Abdulazeez, A. M., Zeebaree, D. Q., Ahmed, F. Y. H. (2021). Predicting university students' performance based on machine learning techniques. In *Proceedings of the IEEE International Conference on Automation, Control and Intelligent Systems* (p. 276–281).

[37] Evangelista, E. D. L., Sy, B. D. (2022). An approach for improved students' performance prediction using homogeneous and heterogeneous ensemble methods. *International Journal of Electrical and Computer Engineering, 12*(5), 5226.

[38] Wu, Y. (2020). Multi-task achievement prediction model based on grade change trend. In *2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence* (p. 1–8).

[39] Yang, Y., Wang, Z., Zhao, S., Wu, J. (2023). An integrated federated learning algorithm for short-term load forecasting. *Electric Power Systems Research, 214*, Article 10883.

[40] Osaka, R., Zhang, T., Zhang, X., Choi, S. S., Chau, T. K., Chow, Y., et al. (2023). Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high PV penetration. *Applied Energy, 333*, Article 120641.

[41] Wang, N., Li, Z. (2023). Short-term power load forecasting based on BES-VMD and CNN-Bi-LSTM method with error correction. *Frontiers in Energy Research, 10*.

[43] Qin, Y., Lu, Z., Liu, H., He, X., Zhang, X., Guo, J. (2023). An improved feature-time transformer encoder-Bi-LSTM for short-term forecasting of user-level integrated energy loads. *Energy and Buildings, 297*, Article 113396.

[44] Wang, X., Yao, Z., Papaefthymiou, M. (2023). A real-time electrical load forecasting and unsupervised anomaly detection framework. *Applied Energy, 330*, Article 120279.

[44] Wang, X., Wang, H., Li, S., Jin, H. (2024). A reinforcement learning-based online learning strategy for real-time short-term load forecasting. *Energy, 305*, Article 132344.

[45] Hussain, S., Hazarika, G. C. (2014). Educational data mining model using rattle. *International Journal of Advanced Computer Science and Applications, 5*(6).

[46] Luo, Y., Han, X., Zhang, C. (2024). Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses. *Asia Pacific Education Review, 25*(2), 267–285.

[47] Thakur, N., Pradhan, S., Han, C. (2022). Investigating the impact of COVID-19 on Online Learning-based Web Behavior. In: Tareq Ahram and Redha Taiar (eds) Human Interaction & Emerging Technologies (IHIET-AI 2022): Artificial Intelligence & Future Applications . AHFE (2022) International Conference. AHFE Open Access, vol 23. AHFE International, USA. https://doi.org/10.54941/ahfe100850.