Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 2, April, 2025

A Comparative Analysis of Machine Learning Models for Predicting EFL Student Language Performance in Smart Learning Environments

Banchakarn Sameephet [1], Wirapong Chansanam [1*], Mahboubeh Rakhshandehroo [2], Chawin Srisawat [1], Kittichai Nilubol [1], Arnon Jannok [1], Bhirawit Satthamnuwong [1], Kornwipa Poonpon [1*]

[1] *Smart Learning Innovation Research Center and Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen, Thailand.*

[2] *Faculty of Culture and Representation, Doshisha Women's College of Liberal Arts, Kyoto, Japan; Center for Multilingual Education, Osaka University, Osaka, Japan.*

Abstract

Integrating smart learning environments into modern education systems opens up significant opportunities to use data analysis techniques to predict students' English language performance. This study aims to evaluate the performance of various machine learning models for predicting English as a foreign language student performance, emphasizing data preprocessing and feature selection. The dataset was gathered from 181 students in eight middle schools in Thailand. The student's data was exported from the Smart Learning Project, which includes data on 14 PISA-like English quizzes covering 27 competencies. The study compares the predictive performance of machine learning models, including Random Forest, Support Vector Regression, AdaBoost, Bayesian Ridge, K-Nearest Neighbors, ElasticNet, XGBoost, Gradient Boosting, and Stacking Ensemble, using MSE, RMSE, MAE, and R^2 metrics. The analysis results indicated that ensemble models, particularly XGBoost and Stacking Ensemble, performed the best in predicting students' English language performance. These models can efficiently capture complex relationships in educational data. Therefore, data preprocessing and feature selection play a significant role in improving model performance. This study highlights the potential of advanced machine learning techniques in educational data analysis. The results can contribute to developing personalized learning strategies and early intervention. It supports an efficient and adaptive education system, advancing smart learning and data-driven instruction.

Keywords:

Machine Learning;
Predictive Modeling;
EFL Student Performance;
Ensemble Methods;
Educational Analytics.

Article History:

Received:	28	November	2024
Revised:	24	February	2025
Accepted:	14	March	2025
Published:	01	April	2025

1- Introduction

There is growing global recognition that a one-size-fits-all approach to education cannot meet the needs of all students or society [1]. In response, countries such as the United States [2], the United Kingdom [3], and China [4] have increasingly embraced personalized learning to address students' diverse needs. Although definitions may differ, personalized learning is generally seen as an educational approach that adapts learning experiences to match each student's strengths, needs, skills, background, and interests [5]. In Thailand, however, personalized learning, especially in the context of learning English as a foreign language (EFL) in smart learning environments, is still limited.

In practice, much of EFL learning in Thailand focuses on both human-human and human-technology interactions. There is a belief that English language learning occurs through interactions supported by technology-enhanced language learning [6]. In practice, the language learning performance of EFL students is often assessed using summative and

* **CONTACT**: wirach@kku.ac.th; korpul@kku.ac.th

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-02-07

formative methods. English teachers play a key role in utilizing these results to evaluate EFL students' current learning progress and predict their future language learning performance. Teachers can use these results as data-driven insights to design personalized English language learning plans for EFL students. The teachers can either manually design personalized learning experiences or use technology to assist them. It is practical for teachers to handle this manually in small classes. However, for large classes or when dealing with extensive data, technology becomes essential to help teachers obtain more accurate insights into student performance. This raises a critical question: What is the most appropriate machine learning model for predicting student performance in learning English as a foreign language in smart learning environments? A comparative analysis is necessary to determine the answer.

The application of machine learning to predict language performance has attracted greater research attention. According to a systematic literature review by Wu et al. [7], which analyzed 83 research articles between 2020 and 2023, ensemble learning produced the best prediction results, achieving an average accuracy rate of 87.67%, followed by the Support Vector Machine (SVM), with an accuracy rate of 84.30%. Important influencing factors used in prediction included demographic, academic, and behavioral factors consistent with Sustainable Development Goal 4 on educational quality. Additionally, Zhao et al. [8] developed a quantitative prediction model by comparing the performance of various machine learning algorithms. The findings suggested that machine learning is an effective tool for identifying educational behaviors and nonlinear relationships between student performance and its influencing factors. Their study highlights the importance of considering multiple influencing factors in model development.

Moreover, Sateesh et al. [9] proposed an ensemble classifier with rule mining utilizing weighted rough set theory and optimizing the weight function with a meta-heuristic algorithm. Test results show that the method outperformed conventional methods with an accuracy of 92.77% and a sensitivity of 94.87%. Furthermore, Çınar & Yılmaz Gündüz [10] tested the performance of several machine learning algorithms, including deep learning and multilayer perceptron, on a dataset from secondary school students in Portugal using 10-fold cross-validation to improve prediction accuracy. Similarly, Şevgin [11] compared the performance of Bagging and Boosting algorithms using TreeNet and Random Forest methods on the dataset from the ABIDE application. The analysis showed that TreeNet outperformed in classification accuracy, sensitivity, F1 score, and AUC value, while Random Forest succeeded in specificity and accuracy.

Handling imbalanced classification in predicting language performance has gained interest in higher education research. Abdul Bujang et al. [12] reviewed the research on imbalanced classification for student grade prediction. The study demonstrated the widespread use of SMOTE (Synthetic Minority Over-Sampling Technique) in determining imbalanced problems and emphasized the importance of hybrid feature selection to enhance prediction performance. Correspondingly, Ye et al. [13] proposed an SA-FEM model using adaptive feature fusion and feature selection to predict online learning performance, which performed better than traditional methods. Li & Yang [14] also developed the XMAMBLSTM algorithm for personalized education resource recommendation by applying deep learning to improve computational efficiency and reduce errors in entity recognition and relationship extraction.

In a related context, Mastrothanasis et al. [15] explored the role of Computational Intelligence (CI) techniques in digital theater performances, highlighting the use of the Flying Fox Optimizer algorithm to form homogeneous student groups and optimize theater dynamics in virtual cultural environments. Moreover, López-García et al. [16] presented a deep learning model based on convolution to address imbalanced classes. This demonstrates its effectiveness in predicting student excellence using features from a large dataset of undergraduate students at the University of Jordan. Finally, Alshamaila et al. [17] presented the application of computational intelligence (CI) techniques in digital theater, using the flying fox optimizer (FFO) algorithm to form student groups and optimize theater dynamics in virtual environments. Malik & Jothimani [18] also tested a deep learning model based on convolution to address imbalanced classes on undergraduate student data at the University of Jordan.

A recent study by Alshamaila et al. [17] presented a model for predicting academic failure that uses the XGBoost algorithm with TOPSIS-based feature extraction and ADASYN oversampling. Malik & Jothimani [18] evaluated the performance of FeatureX in identifying influential predictors and enhancing predictive accuracy to support at-risk students and reduce dropout rates. Integrating advanced machine learning models into educational analytics has revolutionized the field, fostering personalized, efficient, and compelling learning experiences. Sghir et al. [19] provided a comprehensive review of the machine and deep learning models used over the past decade to predict academic outcomes, emphasizing the critical role of predictive modeling in learning analytics. Similarly, Ersozlu et al. [20] highlight various machine learning methods applied to educational data, demonstrating their effectiveness in personalized learning and adaptive assessment. Predictive models are key in forecasting student performance and identifying at-risk students [21]. In line with this, Sajja et al. [22] leverage OpenAI's GPT-4 model for learning analytics tools to assess engagement and track learning progression, underscoring AI's potential in data-driven pedagogical decisions. Considering motivational attributes, Orji & Vassileva [23] developed models incorporating intrinsic and extrinsic motivation to predict study strategies and performance.

The evolution of machine learning in educational research is further captured by bibliometric analyses [24], offering insights into emerging trends and future directions. Brdnik et al. [25] enhance self-regulated learning by aligning learner expectations with performance through predictive analytics. Additionally, deep learning techniques address global challenges in predicting student outcomes in online education [26]. AI models also offer significant potential for improving educational outcomes in developing regions, particularly Latin America [27]. These studies reflect the potential of machine learning to advance inclusive education by predicting student performance, identifying at-risk students, and providing timely intervention. The literature review shows significant progress in applying machine learning in educational contexts, particularly in predicting language performance and identifying at-risk students. The findings highlighted the superior performance of ensemble learning, with its predictive success dependent on incorporating influential factors. While machine learning has advanced language performance prediction, several gaps remain. Current models lack real-time adaptive interventions and often do not utilize multimodal data sources. Issues of imbalanced datasets persist, requiring more sophisticated handling techniques beyond traditional methods to enhance model transparency and scalability across diverse educational contexts.

This study will evaluate machine learning models' performance in smart learning environments to demonstrate their potential for improving English language education, particularly their ability to identify at-risk students to provide timely intervention. Analyzing each model's strengths and limitations allows for selecting tools appropriate to different language learning contexts, leading to the development of more effective and inclusive language learning environments as a cornerstone of raising the overall quality of education.

2- Personalized English Language Learning

Personalized learning is defined by Akyuz [28] as "an educational approach that aims to customize learning for each student's strengths, needs, skills, and interests." Personalized learning involves adapting to a student's unique blend of goals, interests, and competencies, with instruction continuously adjusted to accommodate these evolving factors [29]. Within this approach, learning is shaped by an individual's interactions, including the transfer of knowledge and skills from others and personal experiences [30]. Personalized learning is often seen as a process where EFL students take control of their own language learning, enhancing engagement by fostering a sense of ownership and pride in their progress. Gunawardena et al. [1] highlighted that "personalized learning is touted to provide opportunities for learners to achieve their full potential while developing a love of learning." Thus, the potential of personalization lies in its ability to move beyond a 'one-size-fits-all' educational model, which can disadvantage some students [31].

Personalized learning is closely connected to related concepts such as individualized instruction, adaptive learning, and customized learning, and it fosters the development of knowledge, perspectives, skills, and understanding of individuals [30]. To individualize instruction, Lee et al. [32] suggested that personalized learning could be a solution for addressing students' individual needs and prior experiences, helping them reach their full potential through tailored instruction [33]. Regarding adaptive learning, teachers adjust and tailor students' learning experiences to meet their individual needs. Consequently, personalized learning can effectively enhance students' motivation, engagement, and comprehension [34]. To customize learning, allowing students to select topics, set goals, and choose materials gives them greater control over their learning process [35]. This approach enhances motivation and engagement by aligning the learning experience with their interests and preferences [36].

In English language learning, personalized learning considers students' language proficiency levels, learning styles, interests, and needs to help them improve their English communication skills. Furthermore, personalized learning can boost students' motivation to learn English, promote self-learning, foster independent and critical thinking, and adapt content to align with their interests, learning styles, and existing knowledge. It is designed to address individual needs and goals by tailoring the learning process to specific student characteristics, such as prior knowledge or motivation [37], [38]. Understanding individual learning styles is essential for effective language learning [36]. EFL students exhibit diverse preferences, strengths, and approaches to acquiring new knowledge. Adapting to and supporting these personalized learning styles can significantly enhance language learning. The VARK model, developed by Fleming [39], classifies learners into four categories: Visual, Auditory, Reading/Writing, and Kinesthetic. Visual learners benefit from visual aids and graphics, auditory learners thrive through listening and discussions, reading/writing learners excel with written materials, and kinesthetic learners engage best with hands-on activities. Identifying these learning styles also helps educators and students adopt the most effective methods for acquiring and retaining language skills [40].

However, it is challenging for English teachers to efficiently analyze their students' English language learning needs based on proficiency levels, learning styles, and interests. To support this process, machine learning can be utilized to predict student performance in learning English as a foreign language. By understanding individual students, English teachers can customize language learning to better align with the students' strengths.

3- Using Digital Technology to Support Personalized English Language Learning

Today, personalized English language learning is becoming a popular trend in technology. Many English teachers use technology in their smart learning environments because it helps them adapt teaching materials and provide more meaningful learning experiences that meet their students' needs. Personalized English learning styles extend to preferences in technology-based learning, allowing students to customize their experiences using online platforms and digital tools, such as interactive apps, video lessons, or collaborative forums [41]. Adaptive language learning platforms further personalize instruction by tailoring content to match students' learning styles and progress [42]. Recognizing individual learning styles is essential for optimizing language learning outcomes, as adapting instruction to align with personal preferences enhances its effectiveness.

Technology is becoming increasingly common in language education, as shown by subfields like AI [43], computer-assisted language learning [44], mobile-assisted language learning [45], and technology-enhanced language learning [46]. Along with a wide range of technologies studied, research has shown positive student attitudes and benefits for language learning [47]. These include improvements in motivation, engagement, and confidence [48], as well as better results in receptive skills like vocabulary, grammar, listening, and reading [49] and productive skills such as speaking and writing [50]. According to Liu & Yu [51], the integration and advancement of technology help students overcome barriers to accessing information and knowledge, as well as the limits of traditional learning environments. Additionally, technology supports personalized English language learning by providing digital platforms that allow students to interact with online materials, their peers, and their teachers.

Language learning can be personalized through the use of technology that matches individual preferences [36]. EFL students can access a variety of tools, such as apps, online courses, and digital resources, which can be tailored to their preferred learning styles. These include interactive activities, multimedia content, and group-based tools [41]. This approach allows EFL students to engage with the language in ways that align with their interests, making the learning experience more enjoyable and motivating. Technology-mediated learning is the use of technology to share information and connect people [6]. Examples of digital tools include blogs, learning management systems, mobile apps, social media, and virtual worlds. This approach is now widely used in online English language learning environments. Technology plays a mediating role when English teachers upload learning materials online for students to access later on their laptops or smartphones. Pane et al. [52] stated that technology creates opportunities to address student differences. Changes in how knowledge and skills are acquired can occur through social media and free access to various learning platforms, enabling students to become more independent and innovative. Additionally, Pane et al. [52] emphasized that learning involves fostering curiosity beyond the classroom, encouraging students to participate actively and contribute more to the learning process.

Technology supports personalized language learning by aligning with individual preferences and providing tools like apps, online courses, and digital resources tailored to various learning styles, such as interactive activities and multimedia content [36, 41]. This approach enhances engagement and motivation by matching students' interests. Technology-mediated learning facilitates information sharing and connectivity through tools like blogs, mobile apps, and social media, widely utilized in online English education [6]. English teachers can upload materials for students to access later, making learning more flexible and accessible. Pane et al. [52] highlighted that technology helps address differences among students by enabling changes in how knowledge and skills are acquired, fostering independence and innovation. They also emphasized that technology encourages curiosity beyond the classroom, promoting active participation and deeper engagement in the learning process.

Before designing digital technology to support personalized English language learning, English teachers must first gather solid information about individual students. This information is essential for creating personalized learning experiences supported by digital technology. This study aims to provide insights into machine learning tools that can help teachers predict EFL students' language performance. This knowledge can then be used to design future personalized English language learning programs.

4- Research Methodology

In this research work, a comprehensive analysis was employed to evaluate the efficacy of machine learning models in predicting student performance in smart learning environments. The study encompasses several stages, from data preprocessing, feature selection, model building, and model evaluation. To acquire a deeper understanding of the potential of each educational institution, models have been developed to cover the educational context in all dimensions. The process of conducting a comparative analysis of machine learning models in smart learning environments is shown in Figure 1.

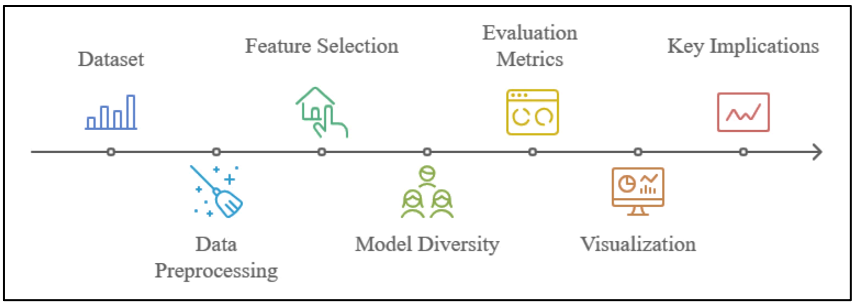

Figure 1. Machine learning process

4-1- Dataset

The study collected a dataset from the Smart English Learning Project, which can be accessed online at smartlearning.kku.ac.th. Thai secondary school students studied English lessons through a digital learning platform (see Appendices 1 and 2). Based on data collected from 181 students in 8 schools, the dataset contains student test results, with 14 PISA-like English quizzes covering 27 competencies. Key features of the dataset are as follows:

- Quiz: Quiz name.
- Ability: Test the ability level.
- User: Unique code for each user.
- Name: Student's name.
- Surname: Student's last name.
- Effort: Effort number for the test.
- Students: Total code for each student.
- Institution: Name of the institution (with several missing values).
- Correct answer: Number of correct answers (with several missing values).
- Highest answer: Highest possible score in the test.
- Score: The actual score the student received (with several missing values).

The dataset contains both student identification data and detailed performance indicators, although several columns, such as 'Institute', 'Correct', and 'Score', contain a large amount of missing data.

Figure 2 shows the results of the correlation analysis between the variables. The strong correlation between 'Correct' and 'Score' indicates consistency between answer accuracy and student performance. Following that, there are significant correlations between 'Maximum' and 'Score' and 'Correct' and 'Maximum', respectively.

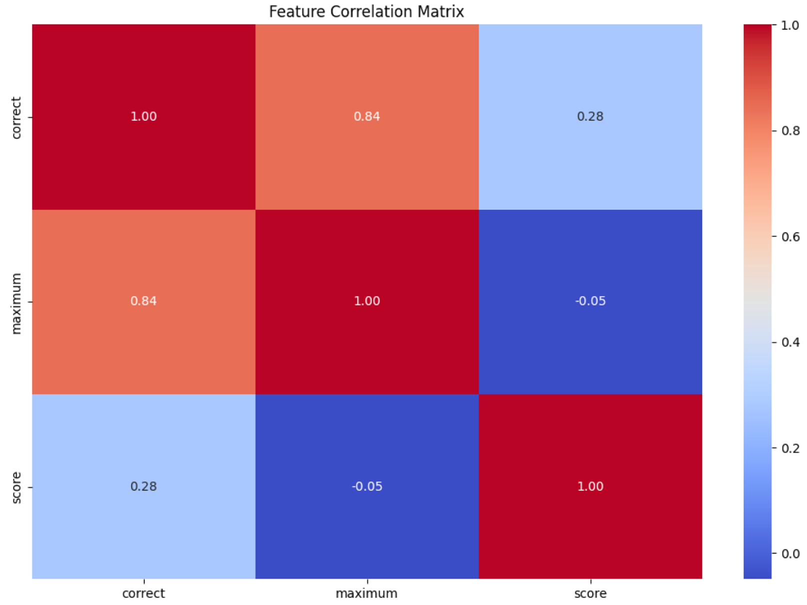

Figure 2. Correlation Matrix of Key Features

The correlation analysis suggests that the number of correct answers and the highest score students can achieve are significant predictors of student learning outcomes. These findings are essential in developing language performance prediction models because these correlations can be used to improve prediction accuracy and teaching and learning planning.

4-2- Data Preprocessing

Data preprocessing is a crucial stage in ensuring the quality and reliability of the dataset. In handling missing values, this study removed rows containing null values in the 'Correct' or 'Score' columns to preserve the integrity of the analysis and prevent errors in the results.

Additionally, when handling qualitative data, particularly in the 'Institute' column containing categorical data, a one-hot encoding technique was applied to convert the data into a format suitable for machine learning models. Data transformation is essential for integrating qualitative data into analytical processes, allowing the most efficient use of each institution's data.

4-3- Feature Selection

Three selected key features for analysis include the 'Correct' column, which reflects accuracy in answers; the 'Maximum' column, which reflects the learner's maximum potential; and the 'Institute' column, which is transformed using one-hot encoding. The selection has created a balance between quantitative metrics reflecting learning performance and institutional factors affecting language performance.

Integrating these features allows for a multidimensional analysis of individual learning outcomes and the influence of the institutional context on language performance. This method provides a comprehensive understanding of the factors influencing student performance, leading to the development of more effective educational approaches.

4-4- Model Diversity

Selecting diverse machine learning algorithms to assess and predict student performance in a smart learning environment is crucial since each algorithm has distinct features and abilities to uncover hidden patterns and relationships in the data.

This study evaluated various regression models, each with strengths and unique data analysis methods. Model diversity allows for comparing the performance and suitability of each model in predictive tasks, leading to the development of a more accurate and reliable prediction system. The assessed models are as follows:

- Random Forest: A collective learning method that creates multiple decision trees and calculates the average of predictions to improve robustness and accuracy. Random Forest uses bootstrap data integration (bagging) to create a diverse subset of training data for each tree [53].

$$(x) = 1/B \sum_{i=1}^{B} f_i(x) \tag{1}$$

where $\hat{f}(x)$ is the Random Forest prediction, $f_i(x)$ is the prediction of the i-th tree, and B is the number of trees.

Random Forest is robust and can handle both classification and regression tasks. It is effective in dealing with data sets that have a combination of categorical and numerical characteristics. This is often found in student language performance data sets. It also handles missing and outliers well and the aggregation nature reduces overfitting.

- Support Vector Regression (SVR): It uses a linear and polynomial kernel to capture the linear and nonlinear relationship between input features and target variables; SVR searches for hyperplanes that maximize margins while accepting partial errors within a tubing that is not susceptible to ε-insensitive tube [54].

$$f(x) = w^T \Phi(x) + b \tag{2}$$

where w is the weight vector, $\Phi(x)$ is the kernel function, and b is the bias term.

SVR is useful for predicting continuous outcomes, such as grades or scores. It can capture linear and nonlinear relationships between features and target variables using different kernels. This flexibility makes it ideal for modeling complex relationships in student language performance data.

- AdaBoost: An aggregation technique that unites weak learners by repeatedly adjusting the weight of incorrectly predicted samples to improve model performance; AdaBoost adjusts the sample weight after adding each weak learner, focusing more on incorrectly classified samples [55].

$$F(x) = \sum^{T}{}_{t=1} \alpha_t h_t(x) \tag{3}$$

where $F(x)$ is the final classifier, $h_t(x)$ are weak learners, and α_t are their weights.

AdaBoost is very useful in improving the performance of weak learners. For student performance information, it can combine many weak models (e.g., simple decision) to create an effective prediction model. This is particularly useful when the initial model is not very accurate when done on its own.

- Bayesian Ridge regression: Bayesian inference is applied to linear regression, allowing the estimation of the parameters of the model; this method introduces a pre-distribution on the parameters of the model to carry out the adjustment [56].

$$p(w|y, X) \propto p(y|X, w)p(w) \tag{4}$$

where $p(w|y, X)$ is the posterior distribution of weights given data X and targets y.

Bayesian Ridge Regression provides probabilistic prediction. This is useful in understanding the predictive uncertainty of the model. It uses normal adjustment, helping to prevent overfitting, especially in data sets with many features and a relatively small number of samples.

- K-Nearest Neighbors (KNN): A non-parametric method that predicts a target value based on the average of k nearest neighbors in a KNN feature area, predicted based on the majority (classification) or average (regression) of k nearest neighbors [57].

$$\hat{f}(x) = (1/k) \sum_{\{i \in Nk(x)\}} y_i \tag{5}$$

where $Nk(x)$ is the neighborhood of x defined by the k closest points.

KNN is a simple and easy-to-use algorithm that works in both classification and regression. This algorithm is especially effective when the relationship between attributes and target variables is non-linear. In student performance data sets, this algorithm can predict student outcomes based on their results perform similar to student tasks.

- ElasticNet: combines the L1 and L2 fines of the Lasso and Ridge methods, making them suitable for data sets with highly correlated properties; ElasticNet combines L1 and L2 standardization to address the co-existence of multiple straight lines and feature selection [58].

$$min(w)\ ||y - Xw||^2 + \alpha[\rho||w||_1 + (1-\rho)/2\ ||w||^2] \tag{6}$$

where α controls overall regularization and ρ balances L_1 and L_2 penalties.

ElasticNet combines the strengths of both Lasso and Ridge regression, making it ideal for data sets with many features, especially when they are interrelated, which is often the case in educational data sets, where different performance indicators can be linked.

- XGBoost: The implementation of efficient gradient boosting that includes normalization and advanced features such as parallel tree generation; XGBoost uses second-order gradient and normalization conditions for more efficient and accurate boosting [59].

$$obj = \sum(i = 1\ to\ n)\ l(y_i, \hat{y}_i) + \sum(k = 1\ to\ K)\ \Omega(fk) \tag{7}$$

where l is the loss function and Ω is the regularization term.

XGBoost is a powerful gradient optimization algorithm known for its high efficiency and effectiveness. This algorithm can handle large data sets with many features and is leveled to avoid overfitting. This makes it ideal for complex tasks such as predicting student performance, which has many factors that may affect the results.

- Gradient Optimization: Model sequentially, with each new model attempting to correct the errors of the previous model; Gradient optimization repeatedly adds weak learners to reduce the distinguishable loss function [60].

$$F_m(x) = F_{m-1}(x) + \gamma_m\, h_m(x) \tag{8}$$

where F_m is the model at iteration m, h_m is the weak learner, and γ_m is the step size.

Gradient Boosting builds sequential models to correct errors of previous models. It is highly flexible and can be used for regression and classification tasks. Its ability to model complex relationships makes it a good candidate for predicting student performance based on a number of factors.

- Stacking Ensemble: Combining multiple machine learning models using metamodels, leveraging the strengths of a variety of models to improve prediction performance; stacking uses predictions from the underlying model as input to the metamodel for final prediction [61]

$$f_{stack}(x) = g(f_1(x), f_2(x), \ldots, f_i(x)) \tag{9}$$

where g is the meta-model and f_i are base models.

Data stacking combines multiple machine learning models to leverage their strengths and improve prediction performance. For measuring student performance, algorithms can integrate different models to capture various aspects of data, making predictions more accurate and stable.

This study highlights the use of various algorithms, including ensemble methods, regression models, probabilistic approaches, and nonparametric methods. This diversity of techniques contributes to capturing student performance data, which is complex and contains many types of data, including numerical, categorical, and missing data.

The selection of the algorithm took into account the balance between a high-bias model and a high-variance model, which is an important factor in developing robust predictive models. Integrating interpretable models (KNN and Bayesian Ridge Regression) with high performance models (XGBoost and Gradient Boosting) allows for understanding predictions while maintaining high levels of accuracy.

The selected algorithm is versatile in handling language performance data sets regarding linear and non-linear relationship analysis, normalization to prevent overfitting, and the ability to handle classification and regression problems. Integrating these algorithms allows for developing comprehensive models that effectively capture the complexity of student performance, leading to more reliable prediction outcomes.

4-5- Evaluation Metrics

Choosing appropriate evaluation metrics for evaluating machine learning models is crucial in predicting student language performance in a smart learning environment. This study defined four key metrics for model evaluation to ensure a comprehensive assessment. Each metric is detailed as follows:

1. *Mean Squared Error (MSE)*

The MSE measures the mean square difference between the predicted and actual values, penalizing larger errors due to squaring [62].

$$MSE = (1/n) \sum(i = 1\ to\ n)\ (y_i - \hat{y}_i)^2 \tag{10}$$

where n is the number of samples, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

MSE is useful when large errors are highly undesirable. In predicting student language performance, significant mispredictions can lead to inappropriate intervention or resource allocation. MSEs will punish these large errors more severely, making it a good choice to avoid significant misjudgment.

2. *Root Mean Squared Error (RMSE)*

RMSE is the square root of MSE. It provides an error measure in the same unit as the target variable, making it more interpretable [63].

$$RMSE = \sqrt{}[(1/n) \sum(i = 1\ to\ n)\ (y_i - \hat{y}_i)^2] \tag{11}$$

RMSE is in the same unit as the target variable (such as test scores), making it easier to interpret for educators and administrators, allowing a clear knowledge of the average forecast error, which is important to understand the practical consequences of model accuracy in an educational context.

3. *Mean Absolute Error (MAE)*

MAE measures the average absolute difference between predicted and actual values. It is less sensitive to outliers compared to MSE and RMSE [23].

$$MAE = (1/n) \sum(i = 1\ to\ n)\ |y_i - \hat{y}_i| \tag{12}$$

The MAE is less sensitive to abnormal values compared to the MSE and RMSE in the educational environment; there may be EFL students with special circumstances that result in abnormal performance; the MAE provides a more stable measure of the model's performance for most students, without being too influenced by these outliers.

4. *Coefficient of Determination (R^2)*

R^2 represents the proportion of variance in the dependent variable that is predictable from the independent variable(s). It ranges from 0 to 1, with 1 indicating perfect prediction [64].

$$R^2 = 1 - [\sum(y_i - \hat{y}_i)^2 / \sum(y_i - \bar{y})^2] \tag{13}$$

where $\bar{y}$ is the mean of actual values, y_i represents the actual values, $\hat{y}_i$ represents the predicted values, i represents the index of each observation.

The R^2 value gives an idea of how much variability the model can explain in EFL students' language performance. This is particularly useful in an educational context as it allows an understanding of the predictive power of the selected characteristics; the high R^2 value suggests that the model can capture a large number of factors affecting students' language performance. This will help to guide other educational strategies and interventions.

Several metrics were used to evaluate the model's predictive performance in this study. The use of different metrics allows for a comprehensive assessment of the model across multiple dimensions. MSE (Mean Squared Error) and RMSE (Root Mean Squared Error) focus on models that can minimize prediction errors. MAE (Mean Absolute Error) helps to gauge the model's error magnitude. R^2 (R-squared) represents the model's ability to explain the variance in student language performance. These metrics provide an overview of the model's performance by addressing the magnitude and direction of errors. The evaluation process involves backtesting to assess metrics' performance and the model's ability to predict unseen data.

The multidimensional assessment ensures that the selected model minimizes prediction errors and provides meaningful insights into the factors influencing students' language performance. This is requisite in the context of a smart learning environment that aims to predict, understand, and continuously improve the learning process.

4-6- Visualization

The results of the model's performance evaluation with each evaluation metric, including MSE, RMSE, MAE and R^2 are presented in bar plots. This visualization provides a clear and objective overview, aiding in the comparison of different models' performance.

Analysts can rapidly and precisely identify each model's strengths and limitations, leading to decisions on the most appropriate model for application in a smart learning environment.

4-7- Key Implications

This study adopts a comprehensive comparative analysis of regression techniques, focusing on leveraging the strengths of ensemble methods to improve prediction accuracy. Combining one-hot encoded institutional data enables the analysis to effectively identify the impact of educational institutions on student performance. Using multiple evaluation metrics helps create a clear understanding of the model's robustness.

This study contributes to laying a foundation for understanding predictive modeling of student language performance in the context of smart learning. Comprehensive comparative analysis of various models and considering institutional impact provides a deeper understanding of the factors affecting student success. Additionally, the findings can be applied to formulate educational policies and develop more effective personalized learning strategies.

5- Results

This section presents the performance evaluation results of each machine learning model in detail. Key evaluation metrics were used in this study, such as the coefficient of determination (R^2), which represents the model's ability to explain the variance of the data, and mean square error (MSE), which reflects the magnitude of prediction error. These metrics are important in evaluating each model's ability to predict student performance and extracting insights from raw data.

Table 1 presents a comparative analysis of different machine learning models for predicting student performance in smart learning environments using several metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). The analyzed models include ensemble methods, support vector machines, optimization algorithms, and regression techniques.

Table 1. Comparative Evaluation of Machine Learning Models for Predicting Student Performance

Model	MSE	RMSE	MAE	R^2
Random Forest	1.790	1.338	0.204	0.997
SVR (Linear)	419.153	20.473	12.418	0.217
SVR (Polynomial)	563.266	23.733	17.151	-0.052
AdaBoost	110.587	10.516	8.352	0.793
Bayesian Ridge	358.883	18.944	12.908	0.330
KNN	14.551	3.815	1.676	0.973
ElasticNet	368.109	19.186	13.284	0.312
XGBoost	0.822	0.907	0.203	0.998
Gradient Boosting	5.469	2.339	1.490	0.990
Lasso	366.726	19.150	13.102	0.315
Ridge	366.804	19.152	13.013	0.315
Huber Regressor	401.274	20.032	12.450	0.250
Stacking Ensemble	1.136	1.066	0.210	0.997

Comparative analysis reveals a significant difference in performance between the models. XGBoost and Stacking Ensemble achieve the highest performance, with the highest R^2 values of 0.998 and 0.997, respectively. Moreover, both models exhibited the lowest error metrics across MSE, RMSE, and MAE, reflecting the high accuracy of their predictions and deriving meaningful conclusions from new data.

Ensemble methods, particularly Random Forest and Gradient Boosting, show outstanding performance, with R^2 values above 0.98 and low errors. These models' success is due to their ability to combine predictions from multiple decision trees, which effectively mitigates overfitting and improves robustness.

In contrast, both linear and polynomial SVR models have limitations in their applicability to this context. Polynomial SVR, particularly shows negative R^2 values, indicating that it performed worse than the horizontal line. K-Nearest Neighbors (KNN) shows moderate performance with an R^2 value of 0.973, outperforming more complex models such as AdaBoost. ElasticNet and Bayesian Ridge also show limitations in their predictions compared to the ensemble methods, with R^2 values of 0.312 and 0.330, respectively.

XGBoost shows the best prediction performance with an R^2 value of 0.998 and low errors across all dimensions: MSE (0.822), RMSE (0.907), and MAE (0.203). XGBoost's success can be attributed to its ability to handle non-linear relationships and interactions between features, as well as its mechanisms to prevent overfitting and efficiently handle missing data. The outstanding success of ensemble methods, particularly XGBoost's leading performance with the lowest MSE and MAE, strongly suggests that future student performance prediction systems should prioritize these sophisticated approaches over simpler models while also considering practical implementation factors such as computational resources and model interpretability requirements.

Stacking Ensemble, which combines multiple base models (Random Forest, XGBoost, and KNN) using Gradient Boosting as a meta-learner, performs similarly to XGBoost with an R^2 value of 0.998, reflecting the effectiveness of leveraging strengths from other models. Similarly, Random Forest performs well with an R^2 value of 0.997, proving the effectiveness of aggregating prediction results from multiple decision trees.

The comparative analysis indicates that although XGBoost shows outstanding performance, a hybrid approach using Stacking Ensemble can achieve similar results, reflecting the potential of integrating strengths from multiple algorithms.

5-1- Feature Importance

Although feature importance scores are not presented explicitly, tree-based models, particularly XGBoost and similar models, are capable of showing these scores. In general, features directly related to student performance, such as 'Correct' scores, are expected to be the most significant predictors. Institutional factors may exert important influences depending on the variance in language performance across institutions.

The analysis results indicate that XGBoost is the best-performing model overall, with the highest R^2 value and lowest error across all dimensions. At the same time, the Stacking Ensemble model shows outstanding performance, reflecting the benefits of combining strengths from multiple models. The combination of Random Forest and Gradient Boosting has proven the effectiveness of the ensemble method in predicting language performance. The findings confirm the potential of advanced machine learning techniques to support educational decision-making, which can lead to the development of personalized learning strategies and early intervention.

In summary, the results demonstrate that the ensemble methods, particularly XGBoost and Stacking Ensemble, yield accurate prediction results and are highly robust. These models not only explain the variance in student performance but also maintain a low error rate, which is an important feature for educational analytics. In contrast, less complex models such as ElasticNet, Bayesian Ridge, and polynomial SVR show limitations in their predictions, reflecting the need for more complex methods to capture the complexity of educational data.

5-2- Visual Analysis

A variety of visual analysis techniques were employed to evaluate model performance and suitability, with an emphasis on box plots and residual plots as the primary analytical tools. Box plots were used to compare the distribution of prediction errors between different models, providing a clear picture of each model's accuracy and variance. It also allows efficient comparison of medians, distributions, and outliers.

As for residual plots, the analysis examines the difference between each model's actual and predicted values. These plots identify residual patterns, indicating the fit between the model and the data. An ideal model shows a random scatter distribution around the center point. The presence of clear patterns or trends in the residuals may indicate areas that need improvement in the model.

Combining these visual analysis tools provides a deeper understanding of model performance by clearly and objectively identifying strengths and weaknesses, leading to more efficient model development.

Figure 3 illustrates the performance of various machine learning models in predicting student scores, showing the relationship between the predicted and actual values. The red dotted line in the diagram represents the ideal situation in which the prediction is perfectly precise. High-performing models such as XGBoost, Gradient Boosting, and Random Forest have prediction points clustered close to the ideal line, reflecting their accuracy and reliability in capturing patterns in the data. In particular, XGBoost shows a close correspondence between the predicted and actual values.

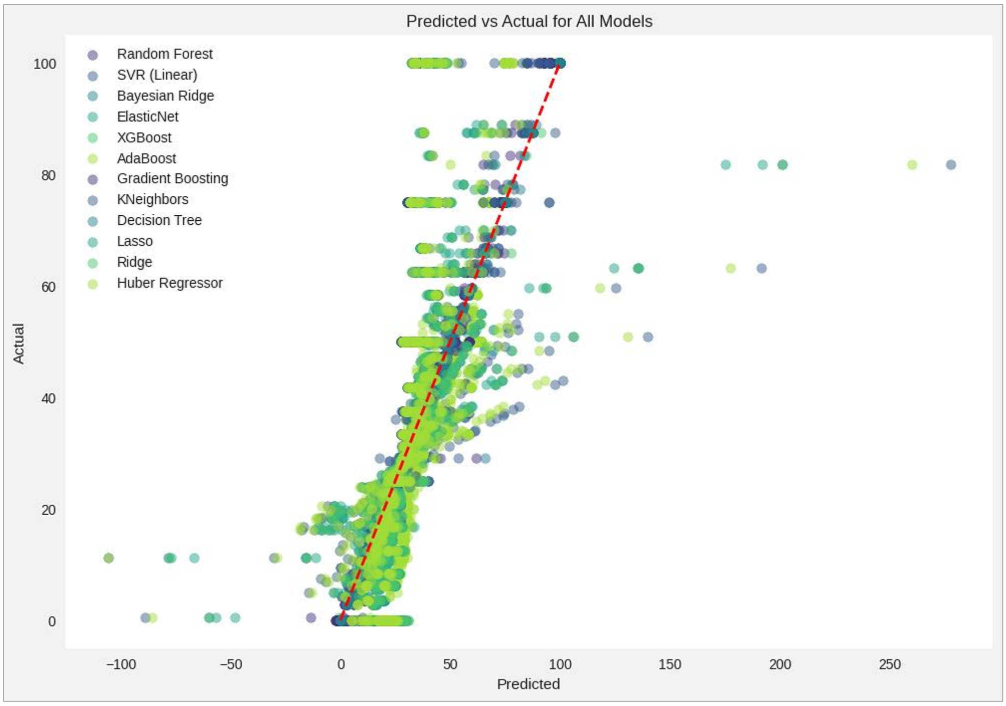

Figure 3. Predicted vs Actual Performance for Various Machine Learning Models

In contrast, polynomial SVR shows limitations, with prediction values scattered away from the ideal line and negative predictions for data with positive actual values, which indicates the model's inappropriateness. KNeighbors and Bayesian Ridge show moderate performance by capturing general trends but lack accuracy compared to the top-performing models.

The overall analysis indicates that ensemble methods, particularly XGBoost and Gradient Boosting, are outstanding in providing accurate and reliable predictions, reflecting their ability to effectively handle the variability in student performance data. The plot also emphasizes the importance of model selection, showing that while some models can produce highly accurate predictions, others struggle to capture the underlying patterns of the data. Ensemble methods have shown an exceptional ability to handle the complexity of educational data, making them a valuable tool for predicting student performance.

Figure 4 depicts the distributions of prediction errors across various machine learning models, including Random Forest, linear and polynomial SVR, AdaBoost, Bayesian Ridge, KNeighbors, ElasticNet, XGBoost, and Gradient Boosting. Each model is represented by a different color for clarity of comparison.

The analysis reveals that the majority of models have a distribution of errors centered at zero, indicating their prediction ability is close to the actual value. In particular, XGBoost, Gradient Boosting, and Random Forest show a high intensity of error at zero, reflecting high prediction accuracy and minimal deviation from the actual values.

In contrast, polynomial SVR shows a wider distribution of errors, with a noticeable spread away from zero, indicating limitations in the prediction and significant deviation from the actual value. Similarly, AdaBoost and Bayesian Ridge show a wider distribution of errors than the top-performing models, reflecting moderate accuracy and less reliability in their predictions.

Overall, the error distribution analysis validates the ensemble methods' superiority, particularly XGBoost and Gradient Boosting, showing the lowest errors and the highest accuracy in predicting language performance. The clustering of errors around zero in these models implies their robustness and efficiency in handling educational data. On the contrary, the wide distribution of errors in other models, particularly polynomial SVR, indicates the need to consider employing more complex methods to improve prediction accuracy.

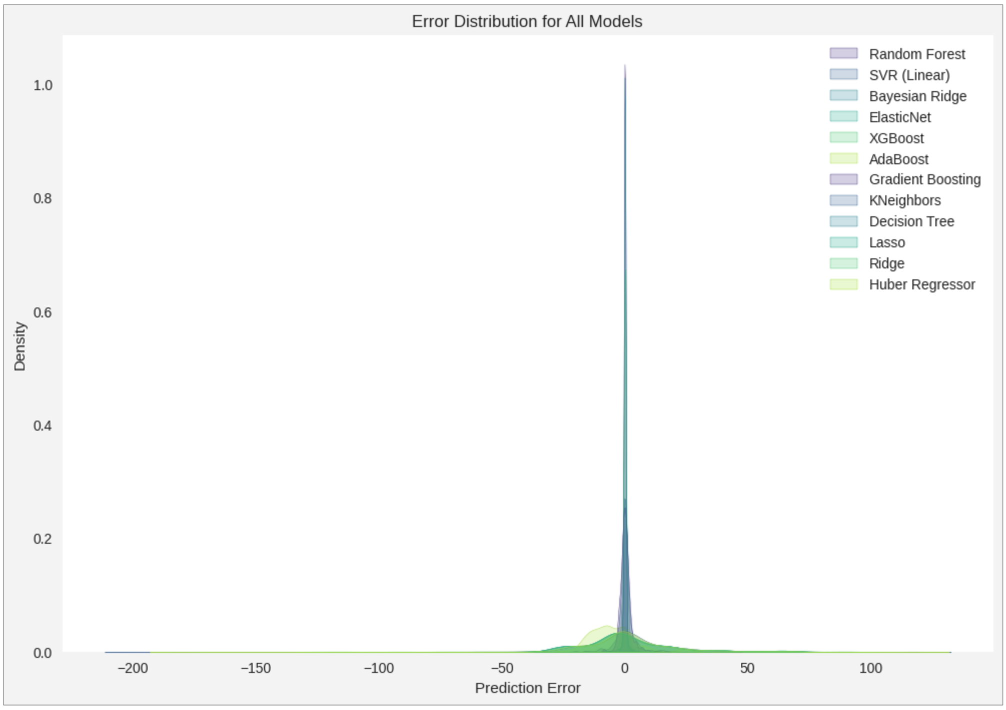

Figure 4. **Error Distribution of Predictions for Various Machine Learning Models**

Figure 5 illustrates the performance of various machine learning models in predicting student language performance based on multiple metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). The ensemble methods—XGBoost, Gradient Boosting, and Stacking Ensemble—consistently demonstrate superior performance, showing the lowest MAE, MSE, and RMSE values and the highest R^2 values. In particular, XGBoost achieves the lowest MAE of 0.203, reflecting its high prediction accuracy. Random Forest and KNeighbors perform moderately well across all metrics, yielding better results than simpler models but falling short of the ensemble methods. In contrast, Polynomial SVR and Bayesian Ridge exhibit significantly higher error values and lower R^2, indicating their limitations in capturing complex data patterns and explaining variability. These results highlight the effectiveness of ensemble methods, especially XGBoost and Gradient Boosting, in minimizing prediction errors and accurately capturing relationships in the data, underscoring their suitability for developing robust language performance prediction systems.

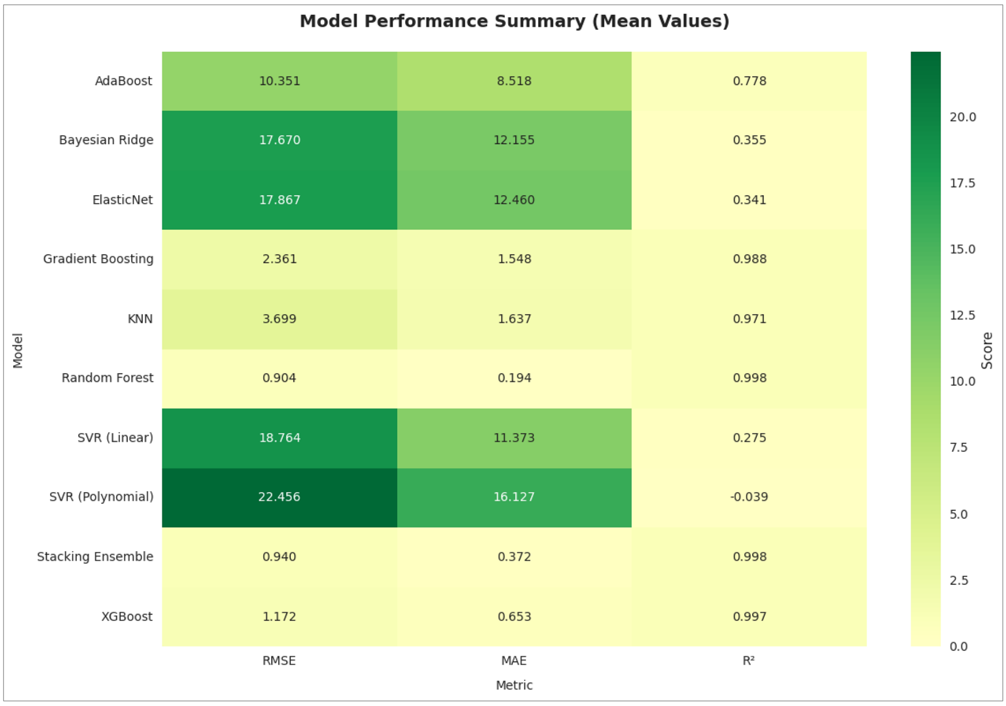

Figure 5. **Model Performance Comparison Summary**

The comparative analysis of machine learning models for student performance prediction reveals a clear hierarchy, with ensemble methods (particularly XGBoost, Random Forest, and Stacking Ensemble) demonstrating exceptional performance through near-perfect R^2 values and minimal error rates, while SVR models struggled significantly. The mid-range performers showed notable variations, with KNN emerging as surprisingly effective (R^2: 0.973) and traditional models achieving only modest success, highlighting the inherently complex nature of educational data. The outstanding results of ensemble methods, especially XGBoost with its leading performance metrics, strongly suggest their prioritization in future educational prediction systems, while considering practical implementation factors such as computational resources and interpretability requirements.

Figure 6 illustrates the residual plot analysis of the machine learning models in predicting student language performance, showing significant differences in the performance of the different models. Ensemble models such as XGBoost and Gradient Boosting show densely clustered residuals around the center, reflecting the model's high prediction accuracy and low errors.

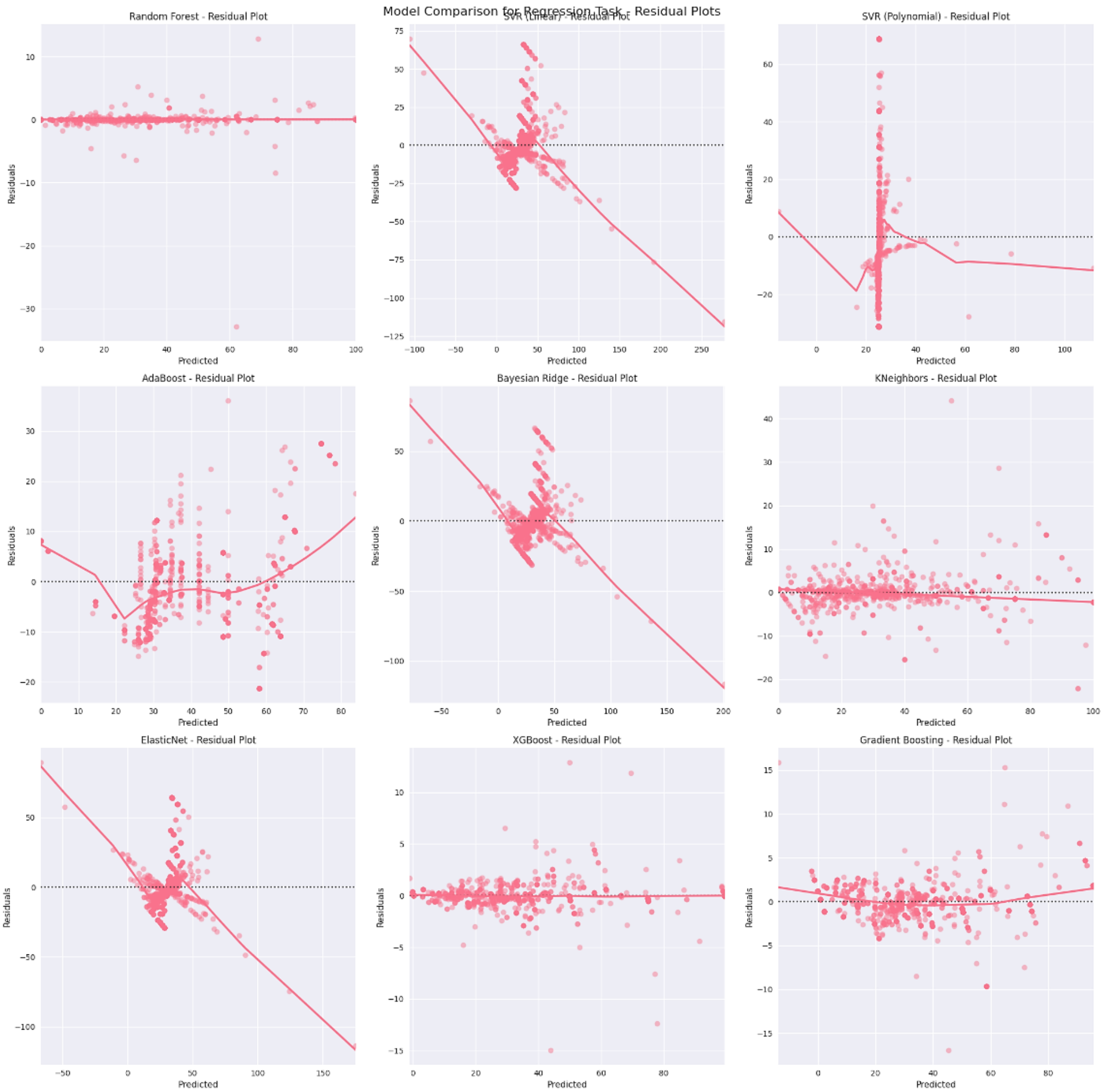

Figure 6. Residual Plots for Various Machine Learning Models

Random Forest and KNeighbors show moderate performance with more dispersed residuals than the ensemble models but still within acceptable limits. In contrast, Polynomial SVR and Bayesian Ridge show a wide distribution of residual values, indicating high errors and inappropriateness of the model in predicting language performance.

Interestingly, AdaBoost and ElasticNet show a systematic pattern of errors, which may indicate bias in the predictions of these models. The results validate the superiority of ensemble models, particularly XGBoost and Gradient Boosting, in predicting student performance with the ability to minimize errors. The findings have significant implications for developing accurate language performance prediction systems and their application in planning preventive education.

Ensemble methods (particularly XGBoost, Random Forest, and Stacking Ensemble) demonstrated remarkable superiority with near-perfect R^2 values and tightly clustered residuals in the comprehensive evaluation of machine learning models for student performance prediction. In contrast, simpler models showed varying degrees of effectiveness, from KNN's surprising robustness to the systematic biases observed in AdaBoost and ElasticNet. The poor performance of SVR models, especially the polynomial variant's negative R^2, combined with their widely scattered residuals, establishes their unsuitability for this specific prediction task. The findings strongly suggest that future educational analytics systems should prioritize ensemble methods while carefully considering the balance between model sophistication and practical implementation constraints, potentially revolutionizing how educational institutions approach student performance prediction and intervention strategies.

Figure 7 illustrates the residual plot analysis of the XGBoost model's performance in predicting student language performance, showing the model's outstanding performance. The residual values are densely clustered around the center line, reflecting the high accuracy of the prediction. Most data points near the red dotted line show minimal deviation between the actual and predicted values.

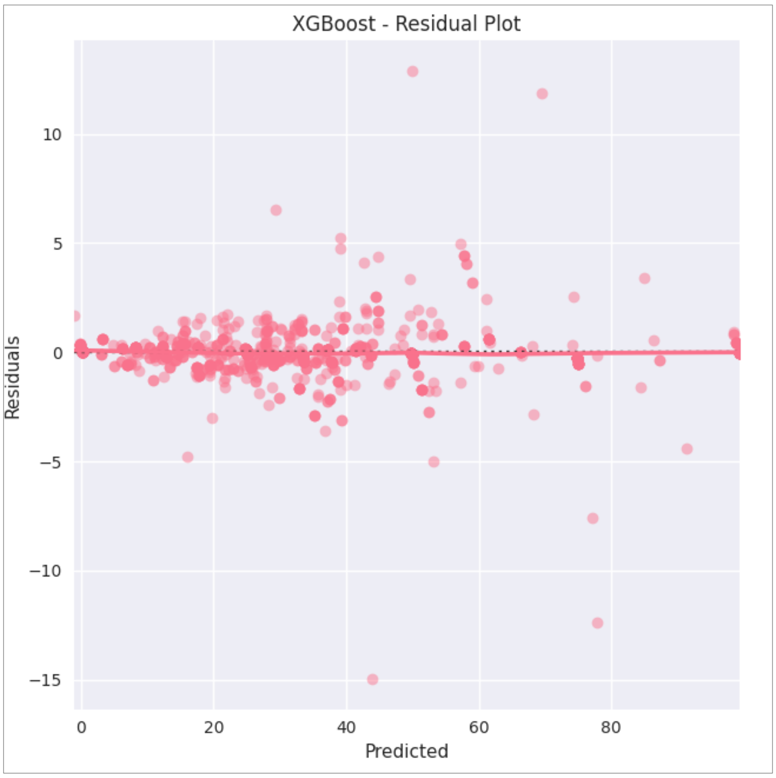

Figure 7. Residual Plot for XGBoost Model

The consistent distribution of residuals over the range of actual values demonstrates that XGBoost can maintain stable prediction accuracy throughout the data set despite a few outliers. It implies that the model can occasionally have discrepancies in its predictions but overall it still produces reliable results.

The results validate the suitability of the XGBoost model for predicting student language performance. The ability to provide highly accurate predictions with low errors makes the model an effective tool for the development of language performance prediction systems and their further application in educational contexts.

The XGBoost model demonstrates exceptional prediction accuracy for student language performance, as evidenced by its superior metrics (MSE: 0.822, R^2: 0.998) and residual plot analysis showing consistently small deviations between predicted and actual values across the performance spectrum. While some outliers exist, the model's ability to maintain stable prediction accuracy across different performance levels, as shown by the even distribution of residuals, suggests it can reliably handle diverse student populations and performance patterns. These results and the model's robust statistical performance establish XGBoost as a highly effective tool for educational prediction systems. However, successful implementation will require careful attention to practical considerations such as regular model maintenance and integration with existing educational frameworks.

Figure 8 illustrates the residual plot analysis of the Gradient Boosting model in predicting student language performance, showing satisfactory performance. Most of the residual values are clustered near the zero line, reflecting the high accuracy of the prediction. The distribution of data points around the red dotted line, with minimal deviation, shows the prediction's reliability.

Figure 8. Residual Plot for Gradient Boosting Model

However, small curves in the residuals, particularly in the high prediction range, indicate minimal systematic error in the model. Despite this limitation, the Gradient Boosting still perform well overall, with most residual values within a narrow range.

The results prove that Gradient Boosting is an effective tool for predicting student language performance with its ability to accurately capture the underlying patterns of the data. Although there are slight inconsistencies in some cases, overall the prediction results are reliable and can be effectively applied to develop language performance prediction systems.

The comprehensive analysis of machine learning models for student performance prediction reveals that ensemble methods, particularly Gradient Boosting, demonstrate exceptional predictive capability with high R^2 values and low error rates. In contrast, residual plot analysis confirms their reliability despite minor systematic errors in higher prediction ranges. The combination of strong performance metrics ($R^2 = 0.990$, MSE = 5.469) and well-distributed residuals around the zero line indicates that Gradient Boosting effectively captures the complex patterns in educational data, though with slight imperfections in specific prediction ranges. The overall results strongly support ensemble methods, especially Gradient Boosting, in developing robust language performance prediction systems, with their minor limitations being far outweighed by their superior predictive accuracy and reliability.

Figure 9 illustrates the residual plot analysis of the Random Forest for predicting student language performance, showing satisfactory performance. Residual values are densely clustered around the center line, reflecting high prediction accuracy. Most data points are close to the red dotted line, indicating low prediction error.

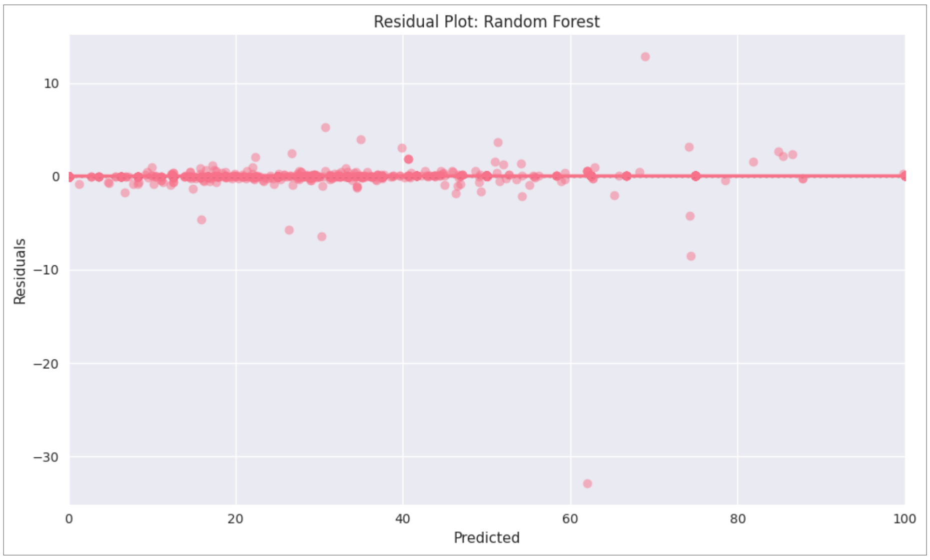

Figure 9. Residual Plot for Random Forest Model

Although some outliers were found, overall residuals remained consistently low throughout the range of predicted values. Random Forest can effectively capture the underlying patterns of the data, making it possible to provide reliable prediction results with low errors.

A small increase in dispersion in the high prediction value range indicates minimal deviation in the prediction. However, the model's overall performance remains stable and reliable, making Random Forest an effective tool for predicting student language performance and can be effectively applied to develop language performance prediction systems.

Table 2 shows the Confusion Matrix of the XGBoost in predicting student language performance, showing outstanding performance. The model achieves a high overall accuracy of 95%, a perfect precision of 1.00 for below-threshold prediction, and a perfect recall of 1.00 for above-threshold prediction.

Table 2. Confusion Matrix for XGBoost Model

	Precision	Recall	F1-Score	Support
Below Threshold	1.00	0.89	0.94	464
Above Threshold	0.90	1.00	0.95	493
Accuracy		0.95	957	
Macro Avg.	0.95	0.94	0.95	957
Weighted Avg.	0.95	0.95	0.95	957

The model's performance is also reflected through high F1-score values of 0.94 for the below-threshold group and 0.95 for the above-threshold group, demonstrating a balance between precision and recall. Additionally, the macro average and weighted average of precision, recall, and F1-score of 0.95 reflect the stability and reliability of the model in classification.

The results prove that XGBoost is highly effective in predicting student language performance. Its ability to accurately and consistently classify language performance levels makes it an effective tool for developing language performance prediction systems and applying them in planning preventive education.

Machine learning models demonstrate varying capabilities in predicting student performance, with ensemble methods (particularly XGBoost and Random Forest) showing exceptional results through multiple evaluation approaches, including near-perfect R^2 values, minimal residuals, and outstanding classification metrics. XGBoost emerges as the slight leader with a 95% accuracy rate and perfect precision/recall in specific categories. At the same time, Random Forest demonstrates excellent prediction stability through its residual analysis, with both models showing remarkable consistency and reliability across different evaluation metrics. The comprehensive evaluation suggests these advanced ensemble methods are highly suitable for developing language performance prediction systems, offering reliable tools for educational planning and early intervention strategies. However, they require more computational resources than simpler models.

Figure 10 depicts the confusion matrix analysis and the importance of the XGBoost features, highlighting several important insights. The confusion matrix revealed strong classification performance. The model correctly identified 412 below-threshold cases (True Negatives) and 493 above-threshold cases (True Positives) while making false predictions, only 52 cases (False Positives), and no cases of False Negatives.

The analysis of feature importance shows that 'Correct' is the most important predictor with a score of 0.533694, followed by 'Maximum', which is the most influential language performance metric with a score of 0.452255. In contrast, institutional features have lower importance scores. The score of 'institution_SKKU' is only 0.013349, reflecting the importance of direct performance metrics over institutional factors.

Empirical evidence demonstrates the superior efficacy of ensemble methodologies, specifically XGBoost and Stacking Ensemble architectures, in delineating intricate correlations within the dataset. Support Vector Regression (SVR) exhibited suboptimal performance metrics, suggesting potential incongruence between the underlying data distribution and the algorithm's fundamental assumptions. While the ensemble approaches demonstrated robust predictive capabilities, subsequent research endeavors should prioritize the systematic optimization of hyperparameters and conduct a more granular investigation into the mechanistic underpinnings of ensemble learning paradigms to enhance predictive accuracy. The comparative analysis reveals that ensemble-based frameworks significantly outperformed traditional SVR implementations in capturing latent patterns, validating the methodological approach while identifying prospective areas for refinement through rigorous hyperparameter optimization and comprehensive exploration of ensemble learning dynamics.

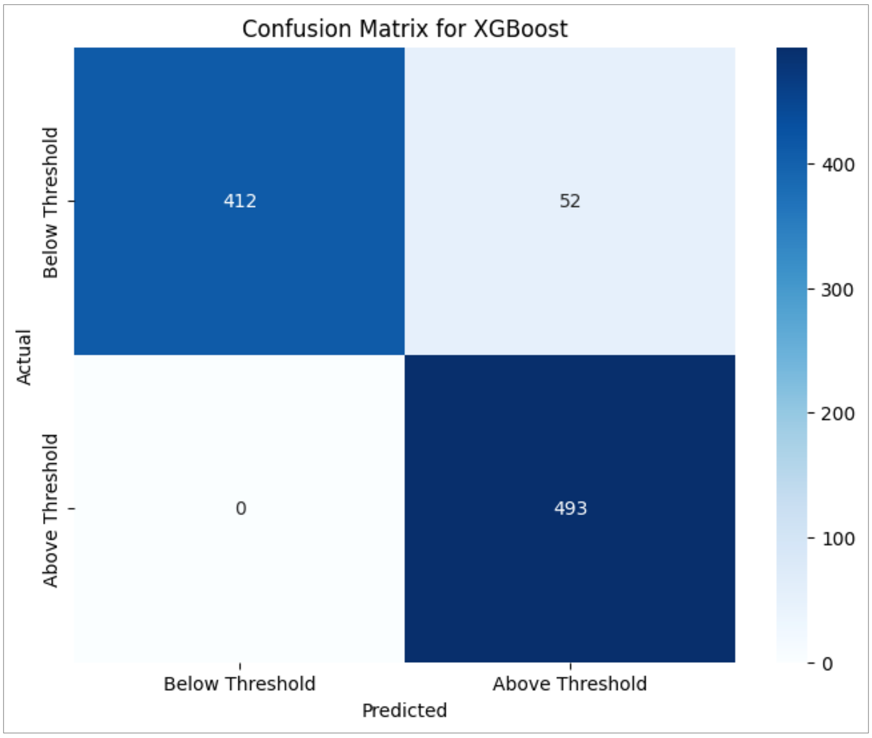

Figure 10. Confusion Matrix for XGBoost Model

6- Discussion

This study conducted a comprehensive evaluation of regression models to predict EFL students' language performance. The analysis results reveal important insights into the various models' performance and suitability. This is consistent with the prior studies by Wu et al. [7] and Zhao et al. [8], ensemble methods, particularly tree-based algorithms, have significant advantages in capturing patterns in data.

Random Forest stood out in the test, with the highest R^2 and lowest error values (MSE, RMSE, MAE). This superior performance is due to the ensemble feature, which mitigates overfitting by averaging predictions from multiple decision trees. Each tree is trained on bootstrap samples with random feature selection. This method improves the generalization of results from diverse data, which is consistent with Zhao et al.'s study [8].

XGBoost and Gradient Boosting produce strong prediction results similar to those of Random Forest. The success of these algorithms lies in a sequential tree-building process where each new tree focuses on correcting the errors made by the previous ones. XGBoost's superior performance compared to traditional Gradient Boosting results from optimized tuning, which includes built-in regularization and advanced tree pruning strategies [8].

In contrast, linear models such as Bayesian Ridge and ElasticNet show moderate performance. Despite its advantages in interpretability and handling linear relationships, its lower performance reflects nonlinear relationships in the data that these models cannot fully capture as stated by Sateesh et al. [9]. The Bayesian Ridge's probabilistic approach and ElasticNet's balanced regularization provided stable, albeit not outstanding, predictions.

Testing with different kernels yields multiple results for Support Vector Regression (SVR). The performance of the linear kernel suggests that simple linear relationships are insufficient for accurate data modeling. The high error rate of polynomial kernels indicates an overfitting problem, reflecting the challenge of selecting appropriate kernels and hyperparameters, as described by Öınar & Yılmaz Gündüz [10].

K-Nearest Neighbors (KNN) performs satisfactorily for nonparametric methods but is not yet comparable to ensemble methods. This reflects the limitations of handling high-dimensional data and its sensitivity to the local structure of the data, as noted by Şevgin [11].

Interestingly, the Stacking Ensemble, which combines predictions from multiple base models, shows strong performance. This is consistent with the study by Abdul Bujang et al. [12], which pointed out that this approach can take advantage of the strengths of various algorithms in capturing diverse aspects of the underlying data structure.

The results highlight the importance of model selection for regression tasks. In particular, the outstanding performance of ensemble methods, both Random Forest and Boosting, reflects the complexity and non-linearity of the relationships in the dataset. However, the practical implementation of these complex models requires consideration of computational costs and limitations in interpreting the results.

The comparison of machine learning models has provided valuable insights for stakeholders in English language teaching and learning contexts. The accurate prediction abilities of models such as Random Forest and XGBoost, according to Wu et al. [7] and Zhao et al. [8], allow teachers and administrators to make informed policy decisions based on aspects of curriculum design, resource allocation, and development of targeted intervention strategies. Additionally, the accuracy of the prediction system helps in identifying at-risk students at an early stage, allowing timely interventions before they escalate, as suggested by Çınar & Yılmaz Gündüz [10].

The ability of ensemble methods to capture complex relationships opens up the possibility of developing individual learning paths. Şevgin [11] pointed out that analyzing the importance of features in high-performing models can lead to more efficient resource allocation. This notion highlights the importance of individualized instructional design. As noted by Lee et al. [32], implementing personalized learning can effectively cater to students' unique needs and backgrounds, enabling them to achieve their full potential through customized teaching approaches [33]. Thus, educational institutions can focus on development in language pedagogy that significantly impact language performance.

In terms of policy, the study results can be used to develop evidence-based education policies, as proposed by Malik and Jothimani [18]. To ensure a clearer language education policy that supports personalized English language learning, it is important for English teachers to adapt and modify language learning experiences to address each student's unique needs. Personalized learning strategies, such as these, can significantly boost students' motivation, involvement, and understanding [34]. Furthermore, when students are allowed to choose topics, set their own learning goals, and select materials, they gain more control over their educational journey [35]. By aligning instruction with individual interests and preferences, this learner-centered approach further increases motivation and engagement [36]. The complexity of prediction models also reflects the need to improve data literacy among educational personnel and support professional development focused on data interpretation and the use of predictive analytics in English language education.

Although the developed prediction model shows interesting potential to support English language learning, its practical application needs to take into account ethical issues, particularly in the areas of data privacy, potential bias, and the responsibility for using predictive analytics in the education system. As López-García et al. [16] proposed, policymakers must develop appropriate guidelines to ensure these technologies are used ethically. Within English language teaching communities, stakeholders—including curriculum designers, teachers, and even parents—should recognize and support the value of personalized English language learning for EFL students. It is important that all parties respect these individualized approaches and act in accordance with clear guidelines. Furthermore, all stakeholders are encouraged to actively participate in the development of tailored policies that address the specific needs of EFL learners.

The model's success in predicting short-term outcomes allows longer-term research into the relationship between early English language performance indicators and long-term educational and career outcomes. The precision of ensemble methods supports the development of adaptive learning systems, as proposed by Smirani et al. [65]. Keser & Aghalarova [66] pointed out that these systems can use real-time data to adjust the difficulty level, pace, and learning activities to suit each learner. Recognizing and understanding individual learning styles is crucial for successful language acquisition [36]. Since students have different preferences, strengths, and ways of learning, adjusting instruction to accommodate these differences can greatly improve their language learning experience. Alshamaila et al. [17] also suggested that model stability across diverse data sets should encourage greater collaboration and data sharing among academic institutions, leading to a deeper and more comprehensive understanding of the factors influencing student success.

The results of this study align with and expand upon previous findings regarding the use of advanced machine learning models in predicting language performance. Alshamaila et al. [17] demonstrated the effectiveness of XGBoost with TOPSIS-based feature extraction and ADASYN oversampling, achieving high accuracy in predicting academic failure. Similarly, Malik & Jothimani [18] highlighted the importance of feature selection techniques in enhancing predictive accuracy for identifying at-risk students. Consistent with these studies, the present study also found ensemble methods like XGBoost and Stacking Ensemble to be the most effective, with the highest R^2 values and the lowest error rates (MSE, RMSE, MAE). The findings support Sghir et al. [19] and Ersozlu et al. [20], who emphasize the critical role of predictive modeling and machine learning techniques in personalized learning and adaptive assessment. While previous studies focused primarily on model accuracy, this study further underscores the importance of data preprocessing and feature selection in improving model performance, suggesting that integrating these methods can lead to more effective and scalable predictive solutions for educational analytics.

The performance of machine learning models may vary across different educational contexts due to variations in curricula, teaching methods, and socio-cultural factors. To improve generalizability, future work should include cross-context validation by testing models on datasets from diverse regions, applying domain adaptation techniques like transfer learning, and incorporating diverse datasets during training to mitigate bias. Additionally, context-specific feature engineering can enhance model relevance by accounting for unique factors in each educational system. These approaches will make models more adaptable, improving their accuracy and fairness in real-world applications.

Data anonymization, de-identification, and encryption should be employed to mitigate data privacy risks and biases in machine learning models, alongside adherence to privacy regulations like GDPR and periodic privacy audits. Differential privacy techniques can further protect sensitive information by adding statistical noise to prevent re-identification. Bias mitigation strategies include ensuring diverse and balanced datasets, using bias-aware algorithms, and conducting regular fairness testing with tools like FairML. Additionally, explainable AI (XAI) methods and ethical reviews involving stakeholders can enhance transparency, accountability, and fairness in model deployment, ensuring ethical and equitable outcomes.

To enhance the interpretability of ensemble models for educators and policymakers, explainable AI techniques like SHAP and LIME can provide clear visualizations of feature contributions. Feature importance scores from models such as Random Forest and XGBoost, combined with interactive dashboards, can simplify complex outputs for practical use. Additionally, surrogate models and stakeholder training sessions can bridge the gap between technical complexity and actionable insights, ensuring that decision-makers fully understand and trust the model's predictions. These approaches enable data-driven decisions while promoting transparency and usability.

Language teacher education programs should include training for English teachers on using machine learning to design personalized English learning, integrating this skill into the national curriculum. By incorporating personalized learning into these programs, English teachers can gain the ability to create technology-enhanced learning experiences tailored to individual needs. Personalized learning involves adapting teaching methods to fit learners' preferences, thereby enhancing their overall learning experience [36]. Teachers can customize digital lessons for visual learners using tools like images, videos, and charts [67]. Auditory learners, on the other hand, benefit from audio resources such as podcasts and conversations [68]. Kinaesthetic learners engage best with hands-on activities like role-playing and interactive games [69]. Providing a variety of materials and tasks supports diverse learning styles, fostering greater engagement and more effective language acquisition. To systematically support English teachers in creating personalized language learning experiences through technology, the VARK model developed by Fleming [39] can be utilized. This model categorizes learners into four types: Visual, Auditory, Reading/Writing, and Kinesthetic. Visual learners benefit most from diagrams and images, auditory learners learn best through listening and speaking activities, reading/writing learners prefer written texts and resources, and kinesthetic learners thrive with hands-on, interactive experiences. By recognizing these different learning styles, teachers and students are better equipped to select strategies that enhance the acquisition and retention of language skills [40].

Future research will focus on longitudinal studies to analyze the lasting impact of early interventions, tracking outcomes like graduation rates and higher education continuation. These studies will help assess the effectiveness of personalized learning strategies over time and inform adaptive educational frameworks for continuous improvement.

Schools with limited budgets can leverage cloud-based platforms, open-source tools, and pre-trained models to reduce costs. Collaborations with academic or industry partners and simpler, resource-efficient algorithms can further facilitate implementation. Investing in English teacher training ensures that these solutions are effectively adopted and maintained, enabling schools to benefit from machine learning without significant financial burdens.

In summary, although machine learning models show great potential for improving English language learning outcomes, their implementation requires balancing technological benefits with ethical considerations and a holistic understanding of the learning process. Integrating these models into educational practice should be performed carefully and their impact continually evaluated to ensure they improve English language performance

7- Conclusion

This study comprehensively evaluates machine learning models to predict EFL students' English language performance in smart learning environments. Various models were analyzed, including Random Forest, Support Vector Regression, AdaBoost, Bayesian Ridge, K-Nearest Neighbors, ElasticNet, XGBoost, Gradient Boosting, and Stacking Ensemble. The analysis provides insight into the potential and limitations of each model in an educational context. The highlight of the study is that it demonstrates the superior performance of ensemble methods. In particular, Random Forest and XGBoost show high prediction accuracy, facilitating the early identification of at-risk students and the development of personalized learning paths. The use of robust cross-validation ensures the performance metrics' reliability and the model's applicability. This highlights the importance of personalized English language learning, as individual preferences now extend into digital environments. EFL students can customize their experiences through online platforms and digital resources such as interactive apps, video lessons, and collaborative forums [41]. Adaptive language learning technologies enhance this personalization by modifying content to match each student's learning style and progress [42]. Recognizing and accommodating these unique preferences is essential for optimizing language learning, as tailoring instruction to individual needs leads to more effective and engaging outcomes.

English teachers can harness the increasing role of technology in language education through tools like AI, computer-assisted, and mobile-assisted language learning. Technology not only boosts students' motivation, engagement, and confidence [47, 48], but also enhances receptive skills (vocabulary, grammar, listening, reading) [49] and productive skills (speaking, writing) [50]. According to Liu & Yu [51], technology helps students overcome access barriers and traditional classroom limitations, while providing platforms for personalized, interactive English learning.

However, the study faces several limitations regarding data quality and availability, ethical issues in terms of privacy and data bias, and challenges in interpreting complex models. Additionally, the focus on short-term predictions without empirical evidence of long-term effects and the high demand for computational resources are significant limitations. Future research should emphasize developing more complex hybrid models and advanced feature engineering to enhance prediction accuracy and robustness. Studies should be expanded to cover a wider range of data sources, and longer-term research should be conducted to evaluate the impact of early intervention and individualized language learning approaches. This will provide an important basis for developing long-term strategies to support success in English language education for the benefit of EFL students in Thailand and beyond.

8- Declarations

8-1- Author Contributions

Conceptualization, B.S[1]. and W.C.; methodology, W.C.; software, W.C.; validation, B.S[1]. and K.P.; formal analysis, W.C.; investigation, C.S., K.N., A.J., B.S[1]., B.S[2]., K.P., and W.C.; resources, C.S., K.N., A.J., B.S[1]., B.S[2]., K.P., and W.C.; data curation, W.C., B.S[1]., C.S., K.N., A.J., B.S[2]., and K.P.; writing—original draft preparation, W.C., B.S[1]., and K.P.; writing—review and editing, W.C., B.S[1]., K.P., M.R., B.S[2]., C.S., K.N., and A.J.; visualization, W.C.; supervision, W.C., B.S[1]., K.P., M.R., B.S[2]., C.S., K.N., and A.J.; project administration, K.P.; funding acquisition, K.P. All authors have read and agreed to the published version of the manuscript.

8-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

8-3- Funding

This work was supported by the Program Management Unit for Human Resources & Institutional Development, Research and Innovation (PMU-B), under the National Higher Education, Science, Research and Innovation Policy Council, Thailand [B05F640229].

8-4- Acknowledgements

The authors would like to express their sincere gratitude to all students who participated in and contributed to this study.

8-5- Institutional Review Board Statement

This research project (Reference No. HE653113) was carefully reviewed and approved by the Khon Kaen University Ethics Committee for Human Research, in accordance with the principles outlined in the Belmont Report and Good Clinical Practice (GCP) guidelines for Social and Behavioral Research (Institutional Review Board Number: IRB00012791).

8-6- Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

8-7- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

9- References

[1] Gunawardena, M., Bishop, P., & Aviruppola, K. (2024). Personalized learning: The simple, the complicated, the complex and the chaotic. Teaching and Teacher Education, 139, 104429. doi:10.1016/j.tate.2023.104429.

[2] Herold, B. (2017). Personalized learning: Modest gains, big challenges, RAND study finds. Education Week, Bethesda, United States.

[3] Kukulska-Hulme, A. (2020). Mobile and personal learning for newcomers to a city. Electronic Journal of Foreign Language Teaching, 17(1), 93–103. doi:10.56040/agkh1717.

[4] Xu, Y. (2021). A Study on the Curriculum Design of Personalized English Teaching in the Context of Internet. Open Journal of Social Sciences, 9(4), 65–71. doi:10.4236/jss.2021.94007.

[5] Bishop, P. A., Downes, J. M., & Farber, K. (2021). Personalized learning in the middle grades: A guide for classroom teachers and school leaders. Harvard Education Press, Cambridge, United States.

[6] Bower, M. (2019). Technology-mediated learning theory. British Journal of Educational Technology, 50(3), 1035-1048. doi:10.1111/bjet.12771.

[7] Wu, M., Subramaniam, G., Zhu, D., Li, C., Ding, H., & Zhang, Y. (2024). Using Machine Learning-based Algorithms to Predict Academic Performance - A Systematic Literature Review. 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), 1–8. doi:10.1109/iciptm59628.2024.10563566.

[8] Zhao, L., Ren, J., Zhang, L., & Zhao, H. (2023). Quantitative Analysis and Prediction of Academic Performance of Students Using Machine Learning. Sustainability (Switzerland), 15(16), 12531. doi:10.3390/su151612531.

[9] Sateesh, N., Srinivasa Rao, P., & Rajya Lakshmi, D. (2023). Optimized ensemble learning-based student's performance prediction with weighted rough set theory enabled feature mining. Concurrency and Computation: Practice and Experience, 35(7), e7601. doi:10.1002/cpe.7601.

[10] Çınar, D., & Yılmaz Gündüz, S. (2024). Classification of Students' Academic Success Using Ensemble Learning and Attribute Selection. Eskişehir Technical University Journal of Science and Technology A - Applied Sciences and Engineering, 25(2), 262–277. doi:10.18038/Estubtda.1394885.

[11] Şevgin, H. (2023). A comparative study of ensemble methods in the field of education: Bagging and Boosting algorithms. International Journal of Assessment Tools in Education, 10(3), 544–562. doi:10.21449/ijate.1167705.

[12] Abdul Bujang, S. D., Selamat, A., Krejcar, O., Mohamed, F., Cheng, L. K., Chiu, P. C., & Fujita, H. (2023). Imbalanced Classification Methods for Student Grade Prediction: A Systematic Literature Review. IEEE Access, 11, 1970–1989. doi:10.1109/ACCESS.2022.3225404.

[13] Ye, M., Sheng, X., Lu, Y., Zhang, G., Chen, H., Jiang, B., Zou, S., & Dai, L. (2022). SA-FEM: Combined Feature Selection and Feature Fusion for Students' Performance Prediction. Sensors, 22(22), 8838. doi:10.3390/s22228838.

[14] Li, S., & Yang, B. (2023). Personalized Education Resource Recommendation Method Based on Deep Learning in Intelligent Educational Robot Environments. International Journal of Information Technologies and Systems Approach, 16(3), 1–15. doi:10.4018/IJITSA.321133.

[15] Mastrothanasis, K., Zervoudakis, K., & Kladaki, M. (2024). An application of Computational Intelligence in group formation for digital drama education. Iran Journal of Computer Science, 1–13. doi:10.1007/s42044-024-00186-9.

[16] López-García, A., Blasco-Blasco, O., Liern-García, M., & Parada-Rico, S. E. (2023). Early detection of students' failure using Machine Learning techniques. Operations Research Perspectives, 11, 100292. doi:10.1016/j.orp.2023.100292.

[17] Alshamaila, Y., Alsawalqah, H., Aljarah, I., Habib, M., Faris, H., Alshraideh, M., & Salih, B. A. (2024). An automatic prediction of students' performance to support the university education system: a deep learning approach. Multimedia Tools and Applications, 83(15), 46369–46396. doi:10.1007/s11042-024-18262-4.

[18] Malik, S., & Jothimani, K. (2024). Enhancing Student Success Prediction with FeatureX: A Fusion Voting Classifier Algorithm with Hybrid Feature Selection. Education and Information Technologies, 29(7), 8741–8791. doi:10.1007/s10639-023-12139-z.

[19] Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). Education and Information Technologies, 28(7), 8299–8333. doi:10.1007/s10639-022-11536-0.

[20] Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. Education and Information Technologies, 1–21. doi:10.1007/s10639-024-12704-0.

[21] Xu, X. (2023). Revolutionizing Education: Advanced Machine Learning Techniques for Precision Recommendation of Top-Quality Instructional Materials. International Journal of Computational Intelligence Systems, 16(1), 179. doi:10.1007/s44196-023-00361-z.

[22] Sajja, R., Sermet, Y., Cwiertny, D., & Demir, I. (2023). Integrating AI and learning analytics for data-driven pedagogical decisions and personalized interventions in education. arXiv preprint, arXiv:2312.09548. doi:10.48550/arXiv.2312.09548.

[23] Orji, F. A., & Vassileva, J. (2022). Machine learning approach for predicting students' academic performance and study strategies based on their motivation. arXiv preprint, arXiv:2210.08186. doi:10.48550/arXiv.2210.08186.

[24] Ayanwale, M. A., Molefi, R. R., & Oyeniran, S. (2024). Analyzing the evolution of machine learning integration in educational research: a bibliometric perspective. Discover Education, 3(1), 47. doi:10.1007/s44217-024-00119-5.

[25] Brdnik, S., Šumak, B., & Podgorelec, V. (2022). Aligning Learners' Expectations and Performance by Learning Analytics Systemwith a Predictive Model. arXiv preprint, arXiv.2211.07729. doi:10.48550/arXiv.2211.07729.

[26] Moubayed, A., Injadat, M., Alhindawi, N., Samara, G., Abuasal, S., & Alazaidah, R. (2023). A Deep Learning Approach Towards Student Performance Prediction in Online Courses: Challenges Based on a Global Perspective. 2023 24th International Arab Conference on Information Technology (ACIT), 1–6. doi:10.1109/acit58888.2023.10453917.

[27] de Souza Zanirato Maia, J., Bueno, A. P. A., & Sato, J. R. (2023). Applications of Artificial Intelligence Models in Educational Analytics and Decision Making: A Systematic Review. World, 4(2), 288–313. doi:10.3390/world4020019.

[28] Akyuz, Y. (2020). Personalized learning in education. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), 69(1), 175-194.

[29] Bulger, M. (2016). Personalized learning: The conversations we're not having. Data and Society, 22(1), 1-29.

[30] Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. Smart Learning Environments, 7(1), 33. doi:10.1186/s40561-020-00140-9.

[31] Fitzgerald, E., Jones, A., Kucirkova, N., & Scanlon, E. (2018). A literature synthesis of personalised technology-enhanced learning: What works and why. Research in Learning Technology, 26. doi:10.25304/rlt.v26.2095.

[32] Lee, D., Huh, Y., Lin, C.-Y., & Reigeluth, C. M. (2018). Technology functions for personalized learning in learner-centered schools. Educational Technology Research and Development, 66(5), 1269–1302. doi:10.1007/s11423-018-9615-9.

[33] Hsieh, C.-W., & Chen, S. Y. (2016). A Cognitive Style Perspective to Handheld Devices: Customization vs. Personalization. The International Review of Research in Open and Distributed Learning, 17(1), 2168. doi:10.19173/irrodl.v17i1.2168.

[34] Pontual Falcão, T., Mendes de Andrade e Peres, F., Sales de Morais, D. C., & da Silva Oliveira, G. (2018). Participatory methodologies to promote student engagement in the development of educational digital games. Computers & Education, 116, 161–175. doi:10.1016/j.compedu.2017.09.006.

[35] Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and Education: The Self-Determination Perspective. Educational Psychologist, 26(3–4), 325–346. doi:10.1080/00461520.1991.9653137.

[36] Al-Mutairi, M. A. (2002). Be Water, My Friend": The Adaptive Approach to English Language Learning. International Journal of Latest Research in Humanities and Social Science (IJLRHSS), 7(5), 20-29.

[37] Sampson, D., & Karagiannidis, C. (2002). Personalised learning: educational, technological and standarisation perspective. Digital Education Review, (4), 24-39.

[38] Tetzlaff, L., Schmiedek, F., & Brod, G. (2021). Developing Personalized Education: A Dynamic Framework. Educational Psychology Review, 33(3), 863–882. doi:10.1007/s10648-020-09570-w.

[39] Fleming, N.D. (2001) Teaching and Learning Styles: VARK Strategies. Neil Fleming, Seattle, United States.

[40] EHRMAN, M., & OXFORD, R. (1990). Adult Language Learning Styles and Strategies in an Intensive Training Setting. The Modern Language Journal, 74(3), 311–327. doi:10.1111/j.1540-4781.1990.tb01069.x.

[41] Lamy, M.-N., & Hampel, R. (2007). Online Communication in Language Learning and Teaching. Palgrave Macmillan UK, London, United Kingdom. doi:10.1057/9780230592681.

[42] Baker, R. S. (2010). Data mining for education: Adaptive learning systems. Handbook of Educational Data Mining, Boca Raton, United States.

[43] Szabó, F., & Szoke, J. (2024). How does generative AI promote autonomy and inclusivity in language teaching? ELT Journal, 78(4), 478–488. doi:10.1093/elt/ccae052.

[44] Cárdenas-Claros, M. S., Dassonvalle, K., Rodríguez-Arias, P., & Cáceres-Ramírez, B. (2024). Considerations for the Design of Pedagogical Tasks in Computer-Based L2 Listening. TESOL Quarterly. doi:10.1002/tesq.3371.

[45] Wu, J. G. (2023). Mobile Assisted Language Learning Across Educational Contexts. Applied Linguistics, 44(6), 1175–1178. doi:10.1093/applin/amac071.

[46] Whitney, N. (2013). Technology Enhanced Language Learning: Connecting Theory and Practice (Oxford Handbooks for Language Teachers). ELT Journal, 68(1), 105–108. doi:10.1093/elt/cct069.

[47] Goksu, I., Ozkaya, E., & Gunduz, A. (2022). The content analysis and bibliometric mapping of CALL journal. Computer Assisted Language Learning, 35(8), 2018–2048. doi:10.1080/09588221.2020.1857409.

[48] Wei, Y. (2022). Toward Technology-Based Education and English as a Foreign Language Motivation: A Review of Literature. Frontiers in Psychology, 13, 870540. doi:10.3389/fpsyg.2022.870540.

[49] Zhang, R., & Zou, D. (2022). A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. Computer Assisted Language Learning, 35(9), 2790–2816. doi:10.1080/09588221.2021.1896555.

[50] Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. Sustainability (Switzerland), 12(2), 524. doi:10.3390/su12020524.

[51] Liu, M. T., & Yu, P. T. (2011). Aberrant learning achievement detection based on person-fit statistics in personalized e-learning systems. Educational Technology & Society, 14(1), 107–120.

[52] Pane, J. F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2015). Continued Progress: Promising Evidence on Personalized Learning. Rand Corporation, Santa Monica, United States. doi:10.7249/rr1365.

[53] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. doi.:10.1023/a:1010933404324

[54] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

[55] Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1), 119–139. doi:10.1006/jcss.1997.1504.

[56] MacKay, D. J. C. (1992). Bayesian Interpolation. Neural Computation, 4(3), 415–447. doi:10.1162/neco.1992.4.3.415.

[57] Fix, E., & Hodges, J. L. (1951). Discriminatory analysis: Nonparametric discrimination: Consistency properties. PsycEXTRA Dataset. American Psychological Association (APA), Washington, D.C., United States. doi:10.1037/e471672008-001.

[58] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67(2), 301–320. doi:10.1111/j.1467-9868.2005.00503.x.

[59] Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. doi:10.1145/2939672.2939785.

[60] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. doi:10.1214/aos/1013203451.

[61] Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259. doi:10.1016/S0893-6080(05)80023-1.

[62] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247–1250. doi:10.5194/gmd-7-1247-2014.

[63] Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79–82. doi:10.3354/cr030079.

[64] Cameron, A. C., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of Econometrics, 77(2), 329–342. doi:10.1016/s0304-4076(96)01818-0.

[65] Smirani, L. K., Yamani, H. A., Menzli, L. J., & Boulahia, J. A. (2022). Using Ensemble Learning Algorithms to Predict Student Failure and Enabling Customized Educational Paths. Scientific Programming, 2022(1), 3805235. doi:10.1155/2022/3805235.

[66] Keser, S. B., & Aghalarova, S. (2022). HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students. Education and Information Technologies, 27(4), 4521–4552. doi:10.1007/s10639-021-10780-0.

[67] Clark, R. C., & Mayer, R. E. (2012). e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. John Wiley & Sons, Hoboken, United States. doi:10.1002/9781118255971.

[68] Field, J. (2010). Listening in the language classroom. ELT journal, 64(3), 331-333. doi:10.1093/elt/ccq026.

[69] Reinders, H., & White, C. (2010). The theory and practice of technology in materials development and task design. English language teaching materials: Theory and practice, 58-80, Cambridge University Press, Cambridge, United Kingdom.

Appendix I

Unit	Topic	Learning Objectives (Can do statements)
1	**My Cool Culture**	Can understand common sentences and phrases related to Isan performances and music (a traditional Northeast Thai show). Can describe Isan performances and music. Can express feelings about Isan performances and music. Can review Isan performances and music.
2	**Social Media & Me**	Can understand common sentences and phrases related to the use of social media. Can read messages on posters related to social media. Can explain how to use social media.
3	**My Ideal Pet**	Can understand common sentences and phrases related to pet care. Can create a poster explaining the ideal pet. Can give a report about the ideal pet.
4	**Making a Shopping List**	Can understand common sentences and phrases related to making a shopping list. Can ask and answer questions about product prices. Can discuss a shopping list with friends. Can give a video presentation about items to purchase.
5	**Superstitions**	Can understand common vocabulary and phrases related to the supernatural. Can explain and provide reasoning about supernatural topics in Thai culture. Can write a short diary about supernatural events.
6	**Get Well Soon**	Can understand common vocabulary and phrases related to health problems. Can describe personal health problems. Can give advice about health problems.
7	**Healthy Me**	Can understand common vocabulary and phrases related to health. Can write a diary about lifestyle and health.
8	**One Tambon One Product (OTOP)**	Can understand common vocabulary and phrases used in presenting local products from their community. Can describe and present local products from their community.

Appendix 2

TIGA Teaching Model	Teaching Objectives	H5P Functions
T (Task)	Explain the main task assigned in each lesson as specified in the textbook.	Image hotspot Find the hotspot
I (Input)	Develop the necessary knowledge required to complete the tasks assigned in each lesson.	*Reading, Vocabulary and Grammar Tasks* Quiz Find the words Image pairing True or false) Drag and drop Memory game *Listening and Speaking Tasks* Dictation Speak the words Answer the questions in Interactive video *Writing tasks* Fill in the blank) Drag the words
G (genre)	Provide examples of text models from different types of communication (genres), explaining the communicative functions of various components and the moves within that text type. Practice language use through sub-tasks and main tasks.	Drag and drop Drag the words Sort the paragraphs
A (Authentic Assessment)	Test learners' knowledge for each lesson by having them read and interpret various types of texts. Self-assess understanding at the end of the lesson.	Quiz True or False Self-evaluation Learning style test