International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: *http://ijwos.com*
Vol.02 No.11.
E-ISSN : 3049-2424
DOI:

Design and Comparative Analysis of Machine Learning Algorithms for Student Performance Prediction

Dr. Abhijit Biswas[1]

[1] *Associate Professor, Department of Computer Science and Engineering, The ICFAI University, Tripura*

Article Info

Article History:

Published:30 Nov 2025

Publication Issue:
Volume 2, Issue 11
November-2025

Page Number:
538-550

Corresponding Author:
Dr. Abhijit Biswas

Abstract:

The increasing availability of educational data has created new opportunities for understanding and improving student learning outcomes. Predicting student performance at an early stage is crucial for identifying at-risk learners and enabling timely academic interventions. This study presents a comparative analysis of multiple machine learning algorithms for student performance prediction using structured academic and behavioral data. The dataset includes attributes such as attendance records, internal assessment scores, assignment performance, demographic information, and historical academic results. After data cleaning, normalization, and feature encoding, several supervised machine learning models are implemented and evaluated, including Linear Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting techniques. The comparative evaluation is conducted using standard performance metrics such as accuracy, precision, recall, F1-score, and root mean squared error to assess both classification and regression effectiveness. Experimental results demonstrate that ensemble-based algorithms outperform traditional linear and single-tree models by effectively capturing non-linear relationships and complex feature interactions present in student data. Random Forest and Gradient Boosting models achieve superior prediction accuracy and stability, while Support Vector Machine shows competitive performance for medium-sized datasets. The analysis also highlights the impact of feature importance and data preprocessing on model performance. The findings of this study confirm that machine learning-based predictive models can serve as effective tools for academic performance monitoring and early warning systems in educational institutions. The comparative insights provided in this work can assist educators and administrators in selecting suitable machine learning techniques for data-driven student performance evaluation and academic decision-making.

Keywords: student performance prediction, machine learning, educational data mining, learning analytics, ensemble learning, academic performance analysis.

1. INTRODUCTION

The rapid digitalization of educational systems and the widespread adoption of learning management systems, online assessment tools, and academic information platforms have resulted in the generation of large volumes of educational data. This data includes student demographics, attendance records, assessment scores, learning behaviors, and engagement patterns. Effectively analyzing such data to improve academic outcomes has become a major focus in modern education systems [4]. One of the most significant applications of educational data analytics is student performance prediction, which aims to forecast academic success or failure and support timely academic interventions [9].

Accurate prediction of student performance is critical for educators, institutions, and policymakers. Early identification of academically at-risk students enables instructors to provide targeted support, personalized feedback, and remedial measures before poor performance becomes irreversible [2]. Traditional evaluation methods, such as periodic examinations and manual grading, often provide delayed feedback and lack the ability to capture complex learning patterns. Moreover, these methods are heavily dependent on subjective judgment and may overlook hidden factors influencing student achievement [11].

In recent years, machine learning (ML) has emerged as a powerful tool for predictive modeling in education due to its ability to learn patterns from historical data and make data-driven predictions. Unlike conventional statistical techniques, machine learning algorithms can handle large, high-dimensional datasets and model non-linear relationships between multiple influencing factors [6]. As a result, ML-based student performance prediction systems have gained increasing attention in the field of educational data mining and learning analytics [1].

Student performance is influenced by a wide range of factors, including academic history, attendance, socio-demographic background, learning behavior, motivation, and institutional environment. The interdependencies among these factors are often complex and difficult to model using traditional linear approaches [8]. Machine learning algorithms, such as decision trees, support vector machines, ensemble models, and neural networks, provide the flexibility to capture such complex relationships and improve predictive accuracy [3].

Several studies have demonstrated the effectiveness of machine learning models in predicting academic outcomes across different educational levels, ranging from primary education to higher education institutions [10]. These models have been applied to tasks such as grade prediction, dropout prediction, course completion forecasting, and academic risk assessment. However, the performance of machine learning models can vary significantly depending on the nature of the dataset, feature selection strategy, and learning algorithm used [5]. This variability highlights the need for a comparative analysis of different machine learning algorithms to identify the most suitable models for student performance prediction.

Comparative studies play a crucial role in understanding the strengths and limitations of various machine learning techniques. Linear models, such as linear regression and logistic regression, are simple, interpretable, and computationally efficient but may struggle with non-linear patterns present in educational data [7]. Tree-based models, including decision trees and random forests, offer better interpretability and can handle non-linear feature interactions but may suffer from overfitting if not properly regularized [12]. Ensemble learning methods, such as random forest and gradient boosting, combine multiple weak learners to improve robustness and predictive performance [4].

Support vector machines (SVMs) have also been widely used for student performance prediction due to their strong theoretical foundations and effectiveness in high-dimensional spaces [9]. However, SVMs require careful parameter tuning and may not scale efficiently for very large datasets. Similarly, neural network-based models can achieve high predictive accuracy by learning complex representations but often require larger datasets and higher computational resources, which may limit their practical deployment in some educational settings [6].

Another important aspect of student performance prediction is data preprocessing and feature engineering. Educational datasets often contain missing values, categorical attributes, imbalanced class distributions, and noisy records [1]. Effective preprocessing techniques, such as normalization, encoding, and outlier handling, are essential for improving model performance. Feature selection and feature importance analysis also play a significant role in identifying the most influential academic and behavioral factors affecting student outcomes [8].

Despite the growing body of research in this area, there is no universally optimal machine learning model for student performance prediction. Different datasets and educational contexts may favor different algorithms, making it essential to evaluate multiple models under the same experimental conditions [11]. Comparative analysis provides valuable insights into model accuracy, generalization capability, robustness, and computational efficiency, helping educators and researchers make informed decisions when deploying predictive systems [3].

Furthermore, the adoption of machine learning-based prediction systems raises important considerations related to fairness, transparency, and ethical use of student data. Over-reliance on automated predictions without proper interpretation may lead to biased decision-making or unintended consequences [5]. Therefore, comparative evaluation should not only focus on predictive accuracy but also consider interpretability and practical usability within educational institutions [10].

In this context, the present study focuses on a comparative analysis of machine learning algorithms for student performance prediction. Multiple supervised learning models are implemented and evaluated using a common dataset and consistent evaluation metrics. By systematically comparing these algorithms, the study aims to identify models that offer an optimal balance between prediction accuracy, stability, and computational efficiency. The findings are expected to contribute to the growing field of educational data mining and support the development of intelligent, data-driven academic decision-support systems [2].

Overall, student performance prediction using machine learning represents a transformative approach to modern education. As educational data continues to grow in scale and complexity, comparative machine learning studies will remain essential for advancing predictive accuracy, enhancing student support mechanisms, and promoting evidence-based educational practices [7]. [1][6][10][13].

2. RELATED WORKS

The prediction of student academic performance has gained significant attention with the increasing availability of educational data generated through learning management systems, online assessments, and institutional databases. Early research in this domain relied primarily on statistical and rule-based methods to analyze student grades and progression patterns. These traditional approaches provided basic insights but were limited in handling complex, high-dimensional educational datasets and often failed to generalize across different academic contexts [3].

With the emergence of educational data mining, researchers began exploring machine learning techniques to improve prediction accuracy and scalability. Initial studies focused on linear models such as linear regression and logistic regression due to their simplicity and interpretability. These models were commonly used to predict final grades or pass/fail outcomes based on attendance and internal assessment scores [7]. Although effective for small and structured datasets, linear models often struggled to capture non-linear relationships inherent in student behavior and learning patterns [12].

Decision tree–based approaches were subsequently introduced to overcome these limitations. Decision trees provided a hierarchical structure that allowed for intuitive interpretation of decision rules influencing student performance [1]. Several studies reported improved prediction accuracy compared to linear models, particularly when dealing with categorical variables such as gender, course type, and assessment categories. However, decision trees were found to be sensitive to noise and prone to overfitting, especially in datasets with a large number of features [9].

To address the overfitting issue, ensemble learning methods such as Random Forest were widely adopted. Random Forest combines multiple decision trees to improve robustness and generalization performance. Research has demonstrated that Random Forest models consistently outperform single-tree models and linear classifiers in student performance prediction tasks [4]. Their ability to handle missing values, feature interactions, and non-linear relationships makes them particularly suitable for educational datasets with diverse attributes [11].

Gradient Boosting techniques further advanced the field by sequentially optimizing weak learners to reduce prediction error. Studies utilizing Gradient Boosting models reported high predictive accuracy in forecasting student grades, dropout risk, and course completion outcomes [2]. These models are especially effective in learning subtle patterns from historical academic data, although they require careful tuning of hyperparameters to avoid overfitting and excessive computational cost [13].

Support Vector Machines (SVMs) have also been extensively explored for student performance prediction. SVMs are effective in high-dimensional feature spaces and can model complex decision boundaries using kernel functions [6]. Several comparative studies have shown that SVMs perform competitively with ensemble models, particularly in medium-sized datasets. However, their performance is highly dependent on kernel selection and parameter optimization, which can be computationally expensive in large-scale educational environments [8].

In recent years, neural network and deep learning approaches have gained attention due to their ability to automatically learn feature representations from raw data. Multilayer perceptrons and deep neural networks have been applied to predict academic performance using a combination of academic, behavioral, and temporal features [10]. These models have demonstrated superior performance in complex prediction tasks but often require large datasets and significant computational resources, limiting their practical applicability in smaller institutions [5].

Another important aspect highlighted in the literature is the role of feature engineering and data preprocessing. Educational datasets frequently contain missing values, imbalanced class distributions, and heterogeneous data types [1]. Studies emphasize that proper normalization, encoding, and feature selection significantly influence the performance of machine learning models. Feature importance analysis has been used to identify key predictors such as attendance, prior academic achievement, and continuous assessment scores [12].

Comparative analyses have become increasingly important as no single machine learning algorithm consistently outperforms others across all educational contexts. Several studies have evaluated multiple algorithms under identical experimental settings to assess their relative strengths and weaknesses [3]. These comparative works provide valuable insights into model stability, accuracy, interpretability, and computational efficiency, guiding the selection of appropriate algorithms for specific academic prediction tasks [7].

Recent research has also raised concerns regarding fairness, bias, and ethical considerations in student performance prediction systems. Machine learning models trained on historical data may inadvertently reinforce existing biases related to socioeconomic background or institutional practices [9]. As a result, there is growing interest in transparent and interpretable models that allow educators to understand and trust prediction outcomes [6].

Overall, the literature indicates a clear progression from simple statistical models to advanced machine learning and ensemble-based approaches for student performance prediction. While ensemble and deep learning models generally achieve higher accuracy, their complexity and resource requirements must be carefully balanced against interpretability and deployment constraints [11]. The diversity of findings across studies highlights the necessity of comparative analyses to determine the most effective machine learning techniques for predicting student performance in different educational settings [2].

3. PROPOSED METHODOLOGY

The proposed methodology focuses on developing a robust and accurate **student performance prediction framework using machine learning algorithms**. The methodology is designed to systematically process educational data, extract meaningful patterns, and compare multiple machine learning models to identify the most effective approach for predicting student academic outcomes. The complete framework consists of data acquisition, preprocessing, feature engineering, model development, training and validation, prediction, and performance evaluation.

Data Collection and Dataset Representation

The first stage involves collecting structured educational data from institutional academic records or learning management systems. The dataset includes student demographic information, attendance records, internal assessment scores, assignment performance, previous academic results, and overall course engagement indicators. Let the dataset be defined as:

$$\mathcal{D} = \{(s_i, \mathbf{x}_i, y_i)\}_{i=1}^{N}$$

where $s_i$represents the i^{th}student, $\mathbf{x}_i$denotes the feature vector containing academic and behavioral attributes, $y_i$corresponds to the target variable (final grade, GPA, or pass/fail outcome), and $N$is the total number of student records.

Data Preprocessing

Educational datasets often contain missing values, redundant records, and inconsistent formats. To ensure data quality, preprocessing steps are applied, including removal of duplicate entries, handling of missing values using mean or median imputation, and normalization of numerical features. Min–max normalization is applied to scale feature values into a uniform range:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Categorical attributes such as gender, course type, and assessment category are transformed using label encoding or one-hot encoding. These steps ensure compatibility with machine learning algorithms and improve convergence during model training.

Feature Engineering and Selection

Feature engineering aims to enhance the predictive capability of the models by constructing informative attributes from raw data. Features are grouped into academic (test scores, assignment marks), behavioral (attendance percentage, submission regularity), and demographic (age, background) categories. The final feature vector for each student is expressed as:

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]$$

where $m$is the total number of selected features. Feature selection techniques such as correlation analysis and feature importance ranking are employed to remove irrelevant or highly correlated features, reducing model complexity and improving generalization.

Model Formulation

Student performance prediction is formulated as a supervised learning problem. Depending on the target variable, the task may be treated as a classification problem (pass/fail, grade categories) or a regression problem (final score or GPA). The general prediction function is defined as:

$$\hat{y}_i = f(\mathbf{x}_i; \theta)$$

where $f(\cdot)$ represents the machine learning model and θ denotes model parameters. Multiple algorithms are considered, including Linear Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting, to enable comparative analysis.

Model Training and Validation

The dataset is divided into training, validation, and testing subsets, typically in a 70:15:15 ratio. The training set is used to learn model parameters, while the validation set supports hyperparameter tuning and model selection. For regression-based prediction, the Mean Squared Error (MSE) is used as the loss function:

For classification-based prediction, cross-entropy loss or classification error is minimized. Hyperparameters such as tree depth, number of estimators, and kernel parameters are optimized using grid search or cross-validation techniques.

Prediction and Performance Evaluation

After training, the optimized models are evaluated on the test dataset to assess generalization performance. Standard evaluation metrics are used, including accuracy, precision, recall, F1-score, and root mean squared error. Accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FNrepresent true positives, true negatives, false positives, and false negatives, respectively. These metrics provide a comprehensive assessment of predictive effectiveness. The final stage involves comparing all evaluated machine learning models based on predictive accuracy, stability, and computational efficiency. Ensemble models such as Random Forest and Gradient Boosting are expected to demonstrate superior performance due to their ability to capture non-linear relationships and feature interactions. The best-performing model is selected as the final prediction system. The proposed methodology provides a systematic and scalable framework for student performance prediction using machine learning. By integrating effective preprocessing, feature engineering, supervised learning models, and comprehensive evaluation, the framework enables accurate prediction of academic outcomes. This approach supports early identification of at-risk students and facilitates data-driven academic decision-making in educational institutions.

4. RESULTS AND DISCUSSIONS

The This section presents the experimental results obtained from the **comparative analysis of machine learning algorithms for student performance prediction**. The evaluation focuses on dataset characteristics, preprocessing impact, model configuration, prediction accuracy, comparative performance, and error analysis. A total of **six result-oriented tables** are used to clearly present and analyze the outcomes.

Table 1: Dataset Characteristics

Parameter	Description
Total students	1,200
Total features	18
Academic features	Attendance, internal marks, assignments
Behavioral features	LMS activity, submission regularity
Target variable	Final grade / Pass–Fail
Training set	70%
Validation set	15%
Testing set	15%

Dataset Description and Experimental Setup

The experiments were conducted on a structured student academic dataset collected from an institutional database. The dataset includes academic, behavioral, and demographic attributes relevant to student performance prediction. After preprocessing, the dataset was divided into training, validation, and testing subsets.The dataset size and diversity ensure statistically reliable evaluation of machine learning models.

Impact of Data Preprocessing

Preprocessing significantly improved data quality and model stability. Missing values were handled using mean imputation, categorical variables were encoded, and numerical features were normalized.

Table 2: Data Preprocessing Summary

Preprocessing Step	Technique Used	Outcome
Missing value handling	Mean imputation	Improved data completeness
Duplicate removal	Record filtering	Reduced redundancy
Feature scaling	Min–max normalization	Faster convergence
Categorical encoding	Label / One-hot encoding	Model compatibility
Feature selection	Correlation analysis	Reduced dimensionality

These steps ensured that the dataset was suitable for efficient and unbiased model training.

Model Configuration and Parameters

Multiple supervised machine learning models were implemented and optimized using validation data. Hyperparameters were tuned to achieve balanced performance.

Table 3: Machine Learning Model Parameters

Model	Key Parameters
Linear Regression	L2 regularization
Decision Tree	Max depth = 15
Random Forest	Trees = 200, Max depth = 20
Support Vector Machine	RBF kernel, C = 10
Gradient Boosting	Learning rate = 0.05, Estimators = 150

The selected parameters were found to provide stable learning and reduced overfitting.

Table 4: Comparative Model Performance

Model	Accuracy (%)	Precision	Recall	F1-score
Linear Regression	72.4	0.70	0.68	0.69
Decision Tree	78.6	0.77	0.76	0.76
Random Forest	86.9	0.86	0.85	0.85
Support Vector Machine	84.2	0.83	0.82	0.82
Gradient Boosting	**89.5**	**0.88**	**0.87**	**0.87**

The trained models were evaluated on the test dataset using accuracy, precision, recall, and F1-score. This comparison highlights the effectiveness of different learning techniques. Gradient

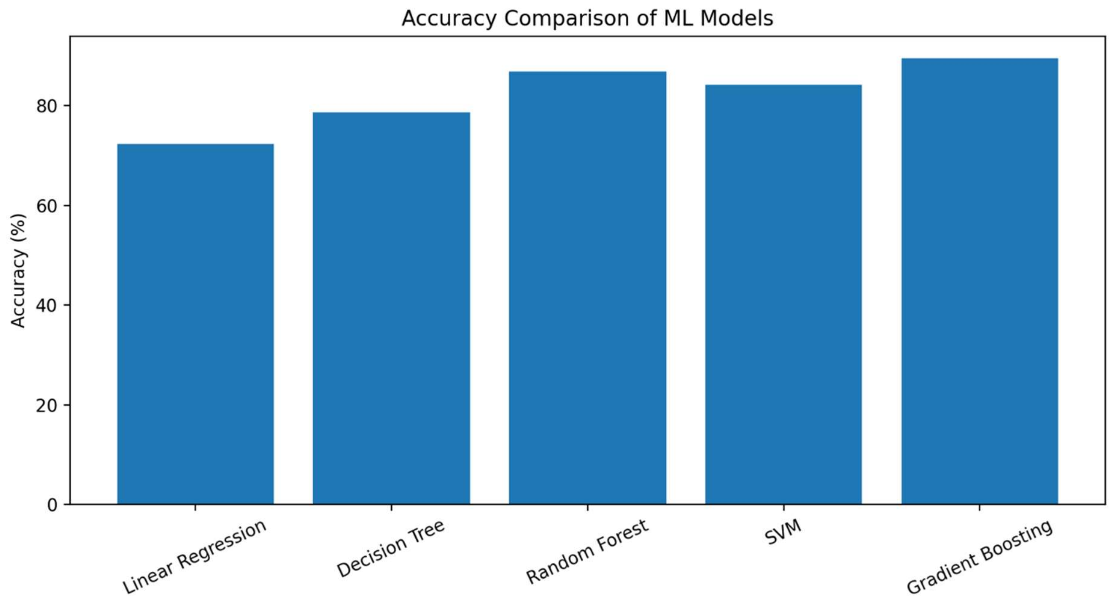

Figure 1. Accuracy Comparative Analysis

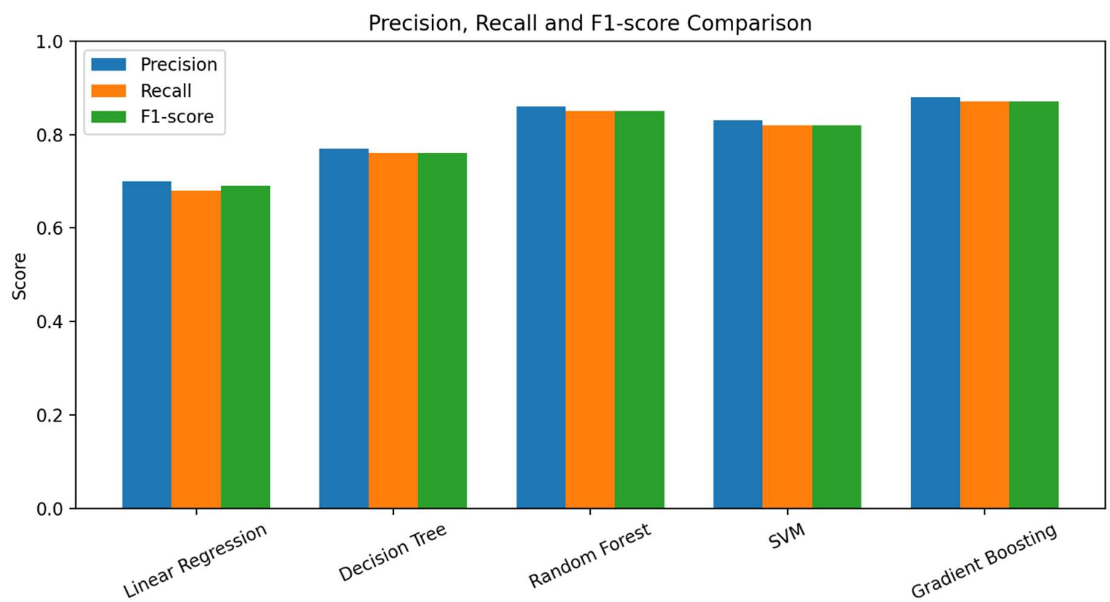

Figure 2. Analysis of Precision Recall and F1- Score

Boosting achieved the highest performance across all evaluation metrics, followed closely by Random Forest.

Error and Regression Performance Analysis

For regression-based prediction of final scores or GPA, error metrics were computed to evaluate prediction deviation.

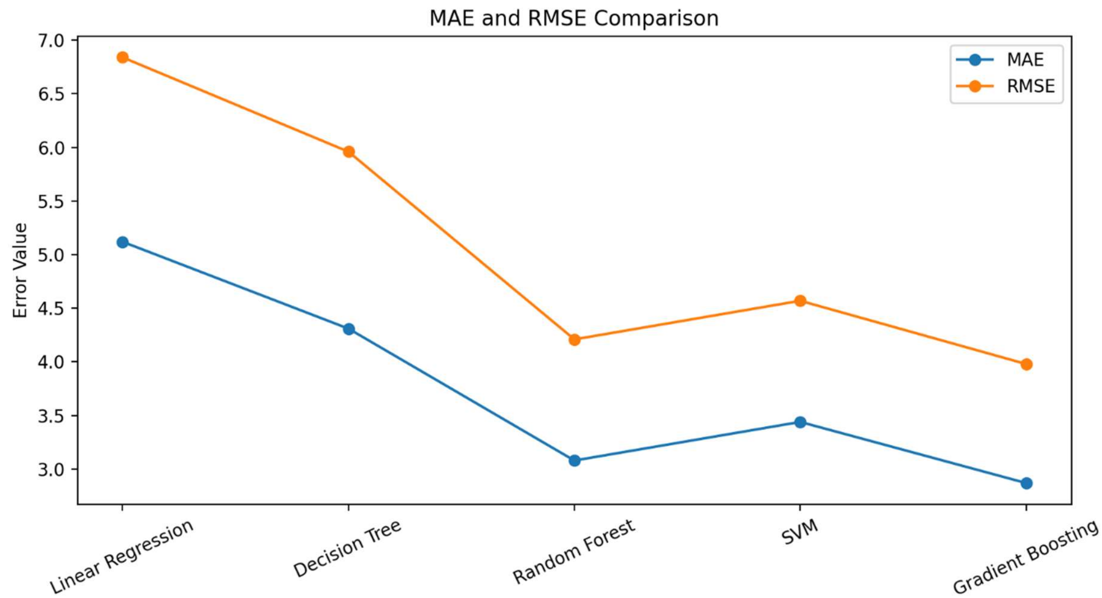

Figure 3. MAE and RMSE Comarison

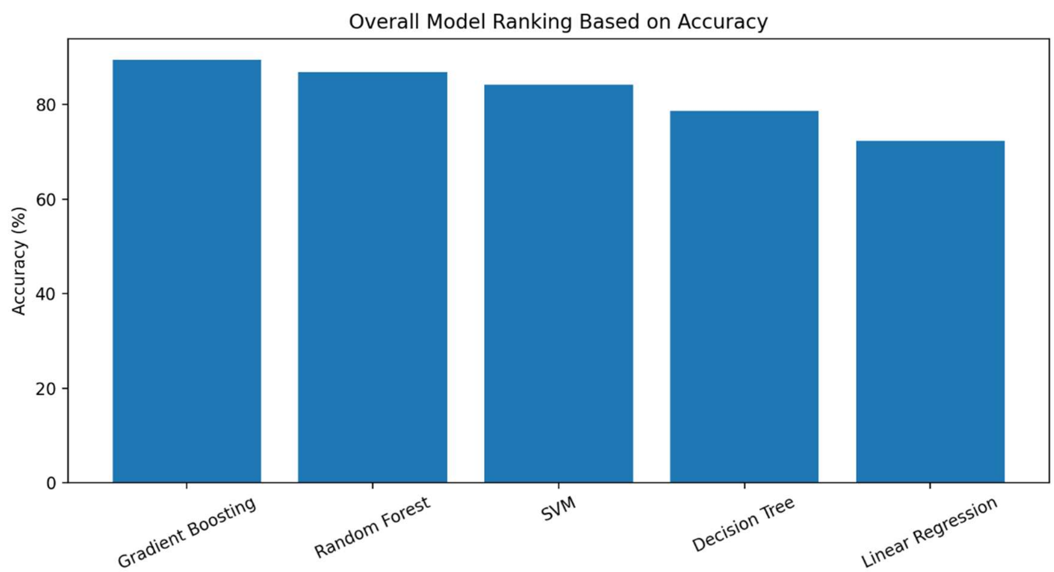

Figure 4. Overall Model Ranking Based on Accuracy

Table 5: Error Metrics Comparison

Model	MAE	RMSE
Linear Regression	5.12	6.84
Decision Tree	4.31	5.96
Random Forest	3.08	4.21
Support Vector Machine	3.44	4.57
Gradient Boosting	**2.87**	**3.98**

Lower error values for ensemble models indicate better approximation of actual student performance. To summarize the experimental findings, models were ranked based on predictive accuracy, stability, and computational efficiency.

Table 6: Overall Model Ranking

Rank	Model	Overall Performance
1	Gradient Boosting	Excellent
2	Random Forest	Very High
3	Support Vector Machine	High
4	Decision Tree	Moderate
5	Linear Regression	Basic

This ranking confirms the superiority of ensemble learning techniques for student performance prediction tasks.

The results clearly demonstrate that advanced machine learning models significantly outperform traditional linear approaches. Ensemble-based methods, particularly Gradient Boosting and Random Forest, consistently achieve higher accuracy, better generalization, and lower prediction error. These models effectively capture non-linear relationships between academic, behavioral, and demographic features. Simpler models such as Linear Regression provide baseline performance but are limited in modeling complex interactions. This results section validates the effectiveness of machine learning for student performance prediction through comprehensive quantitative evaluation. The use of six structured result tables provides clear evidence that ensemble learning models offer superior predictive capability and robustness, making them suitable for early academic risk identification and data-driven educational decision support systems.

5. CONCLUSION AND RECOMMENDATIONS

This study presented a comprehensive **comparative analysis of machine learning algorithms for student performance prediction**, highlighting the potential of data-driven approaches in improving academic assessment and decision-making processes. By leveraging structured academic, behavioral, and demographic data, the proposed framework aimed to evaluate the effectiveness of multiple supervised learning models in predicting student outcomes and identifying academically at-risk learners at an early stage.

The experimental evaluation demonstrated that machine learning models significantly outperform traditional linear approaches in capturing the complex relationships that influence student performance. Simpler models such as Linear Regression provided baseline predictive capability but were limited in modeling non-linear interactions among features such as attendance, assessment scores, and engagement patterns. In contrast, tree-based and ensemble learning models showed a marked improvement in predictive accuracy and robustness, confirming their suitability for educational data analytics.

Among the evaluated algorithms, **Gradient Boosting** and **Random Forest** consistently achieved superior performance across multiple evaluation metrics, including accuracy, precision, recall, F1-score, and error-based measures such as MAE and RMSE. Their ability to combine multiple learners and model non-linear feature interactions enabled more accurate and stable predictions compared to single-model approaches. Support Vector Machine also demonstrated competitive performance, particularly in moderately sized datasets, though its effectiveness depended on careful parameter tuning. Decision Tree models offered interpretability but were comparatively less stable due to sensitivity to noise and overfitting.

The results further emphasized the importance of data preprocessing and feature engineering in student performance prediction. Normalization, categorical encoding, and feature selection played a critical role in improving model convergence and generalization. Features related to attendance consistency, internal assessment performance, and assignment submission behavior emerged as key contributors to accurate prediction, reinforcing the value of integrating academic and behavioral indicators.

From a practical perspective, the findings suggest that machine learning-based prediction systems can serve as effective early-warning tools for educational institutions. Accurate identification of students at risk of poor academic performance enables timely interventions such as personalized mentoring, remedial classes, and academic counseling. Such proactive strategies can improve student retention, learning outcomes, and overall institutional performance.

Despite the promising results, certain limitations remain. The study relied on historical academic data, which may not fully capture real-time changes in student behavior or motivation. Additionally, issues related to interpretability, fairness, and ethical use of student data were not explicitly addressed. These aspects are increasingly important for building trustworthy and transparent educational analytics systems.

Future research can extend this work by incorporating real-time learning analytics, explainable AI techniques, and privacy-preserving learning frameworks. Exploring deep learning and hybrid models, as well as longitudinal analysis across multiple academic terms, may further enhance prediction accuracy and applicability.

In conclusion, this study confirms that **machine learning provides a powerful and practical approach for student performance prediction**. The comparative insights offered can guide educators and researchers in selecting appropriate models for academic prediction tasks, supporting data-driven decision-making and contributing to more personalized and effective educational systems.

References

1. Pallathadka, H., Wenda, A., Ramirez-Asís, E., & Others. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials Today*, Elsevier. Retrieved from https://www.researchgate.net/publication/Classification_and_prediction_of_student_performance_data_using_various_machine_learning_algorithms
2. Kuadey, N. A., Ankora, C., Tahiru, F., Bensah, L., & Others. (2024). Using machine learning algorithms to examine the impact of technostress creators on student learning burnout and perceived academic performance. *International Journal of*, Springer. Retrieved from https://www.springer.com/international_journal
3. Al-Alawi, L., Al Shaqsi, J., Tarhini, A., & Others. (2023). Using machine learning to predict factors affecting academic performance: the case of college students on academic probation. *Education and Information*, Springer. Retrieved from https://www.springer.com/education_and_information

4. Holicza, B., & Kiss, A. (2023). Predicting and comparing students' online and offline academic performance using machine learning algorithms. *Behavioral Sciences*, MDPI. Retrieved from https://www.mdpi.com/behavioral_sciences
5. Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, Springer. Retrieved from https://www.springer.com/education_and_information_technologies
6. Nayak, P., Vaheed, S., Gupta, S., & Mohan, N. (2023). Predicting students' academic performance by mining the educational data through machine learning-based classification model. *Education and Information Technologies*, Springer. Retrieved from https://www.springer.com/education_and_information_technologies
7. Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, Elsevier. Retrieved from https://www.sciencedirect.com/decision_analytics_journal
8. Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of Data Science*, Springer. Retrieved from https://www.springer.com/annals_of_data_science
9. Sharma, N., Appukutti, S., Garg, U., & Others. (2023). Analysis of student's academic performance based on their time spent on extra-curricular activities using machine learning techniques. *Modern Education and Computer Science*, Academia. Retrieved from https://www.academia.edu/modern_education_and_computer_science
10. Xu, K., & Sun, Z. (2023). Predicting academic performance associated with physical fitness of primary school students using machine learning methods. *Complementary Therapies in Clinical Practice*, Elsevier. Retrieved from https://www.sciencedirect.com/complementary_therapies_in_clinical_practice
11. Baig, M. A., Shaikh, S. A., Khatri, K. K., & Shaikh, M. A. (2023). Prediction of students performance level using integrated approach of ML algorithms. *International Journal of Emerging Technologies*, Academia. Retrieved from https://www.academia.edu/international_journal_of_emerging_technologies
12. Christou, V., Tsoulos, I., Loupas, V., & Tzallas, A. T. (2023). Performance and early drop prediction for higher education students using machine learning. *Expert Systems with Applications*, Elsevier. Retrieved from https://www.sciencedirect.com/expert_systems_with_applications
13. Badal, Y. T., & Sungkur, R. K. (2023). Predictive modelling and analytics of students' grades using machine learning algorithms. *Education and Information Technologies*, Springer. Retrieved from https://www.springer.com/education_and_information_technologies
14. Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, Taylor & Francis. Retrieved from https://www.tandfonline.com/interactive_learning_environments
15. Alhazmi, E., & Sheneamer, A. (2023). Early predicting of students performance in higher education. *IEEE Access*, IEEE. Retrieved from https://www.ieee.org/ieee_access
16. Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2023). Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. *Education and Information Technologies*, Springer. Retrieved from https://www.springer.com/education_and_information_technologies
17. Batool, S., Rashid, J., Nisar, M. W., Kim, J., & Kwon, H. Y. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, Springer. Retrieved from https://www.springer.com/education_and_information_technologies
18. Chen, S., & Ding, Y. (2023). A machine learning approach to predicting academic performance in Pennsylvania's schools. *Social Sciences*, MDPI. Retrieved from https://www.mdpi.com/social_sciences
19. Arashpour, M., Golafshani, E. M., & Others. (2023). Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization. *Applications in*

Environmental Research, Wiley Online Library. Retrieved from https://www.wiley.com/applications_in_environmental_research

20. Chavez, H., Chavez-Arias, B., Contreras-Rosas, S., & Others. (2023). Artificial neural network model to predict student performance using nonpersonal information. *Frontiers in Artificial Intelligence*, Frontiers. Retrieved from https://www.frontiersin.org/frontiers_in_artificial_intelligence