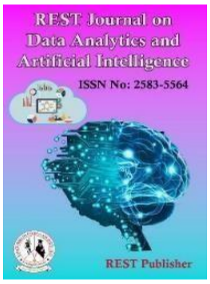
REST Journal on Data Analytics and Artificial Intelligence

Vol: 4(4), December 2025

REST Publisher; ISSN: 2583-5564

Website: http://restpublisher.com/journals/jdaai/

DOI: https://doi.org/10.46632/jdaai/4/4/12



Predicting Academic Performance Through AI Use in Schools: A Random Forest Regression Approach

Vinod Kumar A

Maharshi Valmiki College of Education, Delhi, India.

Corresponding Author Email: tagvinod@gmail.com

Abstract: *This study, using a comprehensive mixed methods approach, examines the relationship between artificial intelligence (AI) use in educational settings and academic performance. Using a random forest regression model, the research examines how daily AI study hours are associated with student achievement in math, science, and English. Analysis of 100 student observations reveals a strong positive association between increased study time and academic scores, with students with high AI use (9+ hours) consistently achieving scores in the 90-100 range, while those with minimal use (below 2 hours) typically achieving scores in the 50-60 range. Descriptive statistics indicate a mean study time of 5.2 hours, with mean scores of 76 for math and English and 69 for science. The random forest models showed exceptional performance on the training data ($R^2 > 0.95$) but showed reduced generalization on the test data, indicating some over fitting tendencies. Despite this, the models maintained reasonable predictive accuracy with experimental R^2 values ranging from 0.71 to 0.89 across subjects. The findings highlight the transformative potential of AI in education, while identifying important challenges including infrastructure limitations, data privacy concerns, and the need for strategic implementation frameworks. The study emphasizes the importance of integrating AI with human-cantered teaching approaches to enhance educational outcomes.*

keywords: *Artificial Intelligence in Education, Random Forest Regression, Educational Performance, Data-Driven Decision Making, Machine Learning, Educational Technology, Predictive Modelling*

1. INTRODUCTION

The impact and practical application of data and artificial intelligence (AI) in educational management. Adopting a mixed-method approach, it combines quantitative analysis with qualitative insights from interviews and focus groups, reflecting the views of teachers, administrators, and students on the use and effectiveness of these technologies. The research seeks to explore how data and AI are being integrated into educational systems, explore their potential benefits, and identify challenges that hinder their success. Findings point to positive outcomes in areas such as governance and curriculum development, but also reveal barriers such as lack of adequate infrastructure and resistance to innovation. Overall, the study emphasizes the importance of strategic investment and human-cantered teaching approaches to unlock the full potential of data and AI in education [1]. As the education sector undergoes rapid transformation, artificial intelligence is positioning itself as a catalyst for profound change, redefining conventional approaches to decision-making. This study presents a comprehensive framework powered by AI, developed to enhance data-informed strategies for improving schools. Using cutting-edge algorithms, this framework helps educators and administrators understand complex patterns in student performance, assess teaching effectiveness, and refine resource allocation. This methodology not only responds to the growing importance of data in education, but also aligns with the broader shift towards intelligent and adaptive learning ecosystems, with the ultimate goal of improving educational outcomes [2]. The integration of artificial intelligence is gaining importance in improving educational delivery, especially in junior high schools and equivalent Islamic institutions such as Madrasah Saniya. This research explores the critical role of AI in personalizing learning experiences, analysing student performance data, and improving administrative processes [3]. The literature review indicates that AI supports adaptive learning by assessing individual student needs and providing data-driven recommendations, while improving operational efficiency through the automation of administrative tasks. Although challenges such as limited infrastructure and insufficient professional training persist, AI has significant potential to revolutionize education. With targeted investments and extensive training efforts, it can contribute to creating a more inclusive and sustainable educational environment [4]. Smart classrooms are advanced learning environments that use data collected from students, teachers, and the surrounding environment to improve educational practices. The analysis of this data enables informed decision-making and improved learning outcomes. This research paper presents an in-depth review of the current applications of artificial intelligence (AI) techniques in analysing such data to gain a deeper understanding of teaching methods [5]. This study reviews the existing literature, establishes a taxonomy to classify AI application areas, assesses the current level of implementation in schools, and identifies barriers to wider adoption. The integration of AI into real primary and secondary smart classroom environments is still in its infancy. The results reveal that many efforts have neglected important aspects such as information security, data privacy, and the measurable impact of AI on student performance [6]. While artificial intelligence holds great promise for transforming education through personalized learning, automated administration, and enhanced student engagement, there is a major challenge in effectively using this technology to improve learning outcomes. Central to this issue is the role of leadership. This study seeks to bridge this gap by examining how different school leadership styles influence the effective use of AI for meaningful educational progress [7]. While proponents of advanced AI systems like Galactic and ChatGPT argue that they will transform education, their initial releases have sparked considerable controversy. Both models have faced criticism for generating seemingly reliable but actually false information and for producing educational content that is virtually indistinguishable from human writing, thereby raising serious concerns about educational integrity and potential fraud [8]. Explores the impact of AI-based tools on student engagement in high school chemistry classrooms. The goal is to provide educators with practical approaches to using AI to design more interactive and effective learning experiences in chemistry and science education overall [9]. The rapid advancement of artificial intelligence is reshaping society, and schools are key centres of this transformation. As institutions responsible for preparing students for the digital age, they must adapt to emerging AI technologies by incorporating 21st century skills. Given the pervasive influence of these tools and their strong appeal to younger generations, such adaptation is not optional but necessary, as noted by Karsenti (2019) [10].

2. NURSING CARE PLANNING

Materials

Study Hours: Study time refers to the total number of hours a student is expected to dedicate to completing all learning and assessment activities within a course or its specific components. This concept is closely tied to the academic credits allocated and serves as a key indicator of the feasibility of the course and the learner's ability to manage the workload.

Math Score: The evolution of mathematics is a gradual process spanning thousands of years and many cultures, rather than the creation of a single individual. Although Archimedes is often hailed as the "father of mathematics" for his significant contributions, its basic foundations were laid by earlier inventors from ancient civilizations, including Sumer, Egypt, Greece, India, and scholars of the Islamic Golden Age.

Science Score: A science score is a numerical indicator of a student's proficiency or achievement in a science subject. It is typically obtained through a combination of standardized tests, exams, or assessments, and laboratory work. This score serves as an important tool for assessing understanding of science concepts, monitoring academic progress, comparing performance among students, and guiding decisions about academic placement and future science-related goals.

English Score: An English score is a numerical measure used to assess a student's English language proficiency and comprehension. It is typically based on assessments of key skills, including reading comprehension, writing, grammar, and sometimes listening and speaking skills. This score serves as a standardized indicator of academic performance, helps track progress, guides placement in academic programs, and helps meet graduation or university admission requirements.

Instructions For Machine Learning

Random Forest Regression: Random Forest is an ensemble learning technique designed to predict continuous values. It uses randomly selected subsets of data and features to construct multiple decision trees before combining their results by averaging the results. This method reduces the potential for over fitting and improves the generalizability of the model. The randomness between trees introduces heterogeneity, which leads to more reliable and accurate predictions than a single tree. Random Forest Regression is well suited to capturing complex, nonlinear relationships and managing high-dimensional datasets. It is resilient to noise and outliers and provides insightful information about the importance of features, revealing the most influential variables. Its reliability, flexibility, and high accuracy make it a popular choice for predictive modelling in fields such as finance, healthcare, and engineering.

3. RESULT AND DISCUSSION

This dataset provides a dataset that examines the relationship between daily AI usage (in hours) and academic performance in three subjects. Each row represents data from an individual student. A clear positive correlation is observed: as study time increases, scores in math, science, and English generally increase. For example, students with higher study time (e.g., 9.73, 9.88) consistently score in the 90s or even 100, while those with lower study time (e.g., below 2) typically score in the 50s and 60s, demonstrating a strong correlation between study time and academic achievement.

TABLE 1. Descriptive Statistics

	Study Hours	Math Score	Science Score	English Score
count	100.00000	100.00000	100.00000	100.00000
mean	5.23210	76.08760	68.65470	76.70110
std	2.67740	13.35070	13.44381	12.27342
min	1.05000	52.45000	44.65000	54.17000
25%	2.73500	64.41500	58.59250	66.10000
50%	5.17500	74.14500	66.69000	75.47500
75%	7.57500	88.02750	77.74250	86.95000
max	9.88000	100.00000	100.00000	100.00000

Table 1 provides a descriptive summary of the dataset. The average student spends approximately 5.2 hours studying, with an average score of 76 for Math and English, and a low score of 69 for science. The standard deviations (~13 for scores, ~2.7 for hours) indicate moderate variation in student performance and study habits. The score ranges approximately from 52–100, confirming that it represents a full range of performance. In particular, the median (50th percentile) and mean values for all subjects are close, indicating a relatively symmetrical distribution of scores without extreme skewness, providing a reliable basis for subsequent predictive modelling.

Effect of Process Parameters

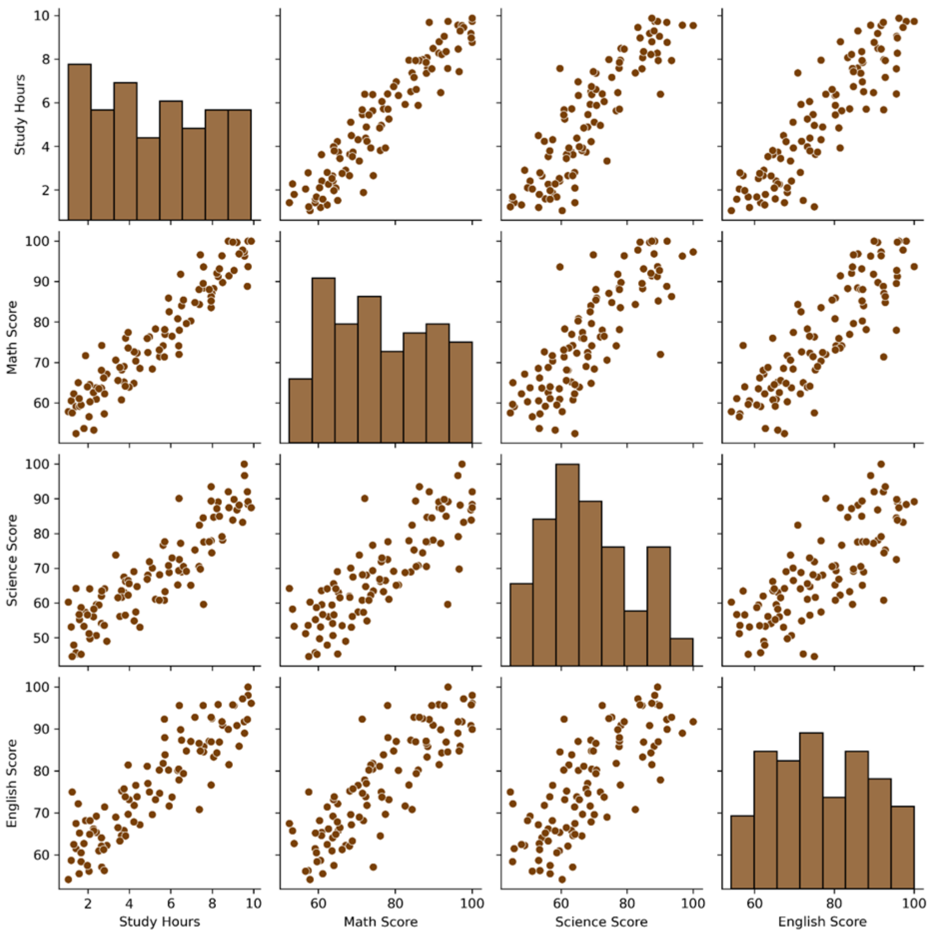


FIGURE 1. Scatter plot of the Data Usage AI in School

Figure 1 presents a scatterplot illustrating the data use of AI in schools. Each point represents an individual observation, highlighting the variations in AI use across contexts. The distribution indicates patterns of adoption, with some clusters showing high engagement and others showing very low engagement, reflecting varying institutional priorities, access, and integration of AI technologies.

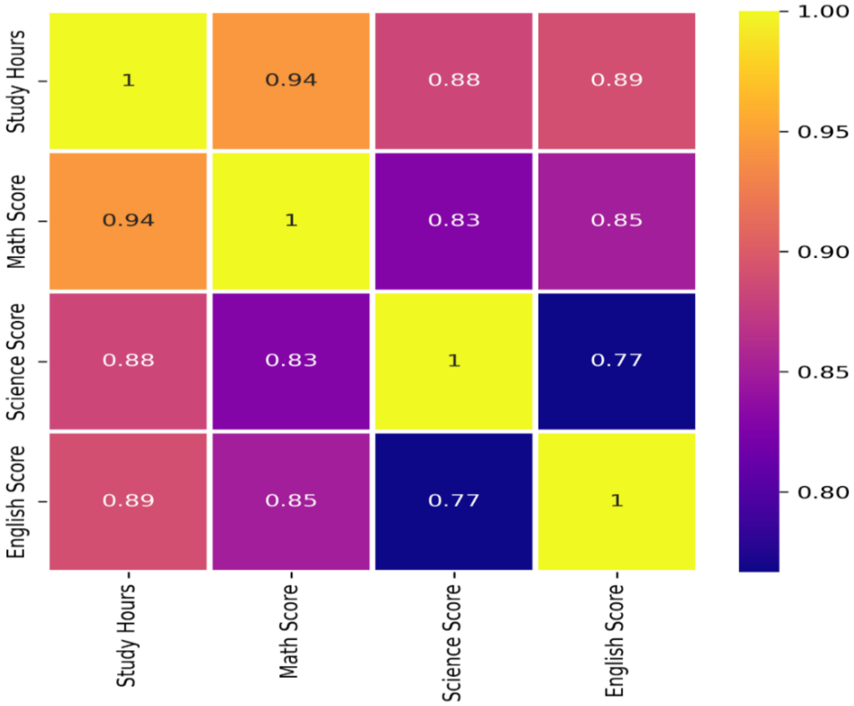


FIGURE 2. Heat map of the relationship between process parameters and outcomes

Figure 2 shows a heat map depicting the relationship between process parameters and outcomes. Colour intensity indicates the strength of the correlation, with darker shades reflecting stronger correlations. This visualization helps identify critical parameters that influence results, revealing trends, biases, and optimization opportunities that are essential to improving performance and achieving desired performance outcomes.

Random Forest Regression (Math Score)

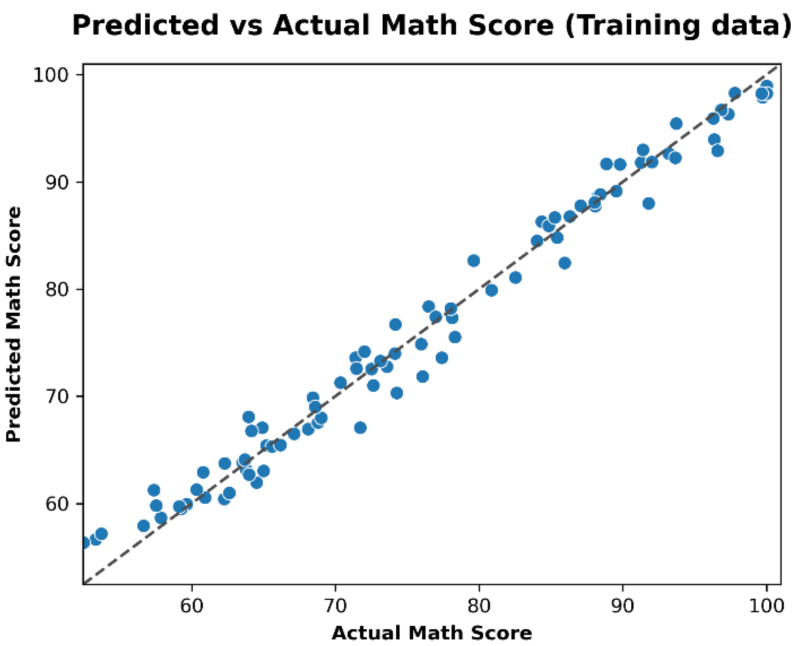


FIGURE 3. Random Forest Regression on Math Score: training data

Figure 3 shows that the Random Forest model demonstrates excellent predictive accuracy for math scores on the training data, with points densely aligned along the diagonal reference line across the entire score range (55-100). The tight clustering indicates that the model has effectively learned the underlying relationships, with minimal prediction errors. Some minor scatter appears in the mid-range scores, but overall, the model shows a strong fit and captures variation across different performance levels well without obvious systematic bias.

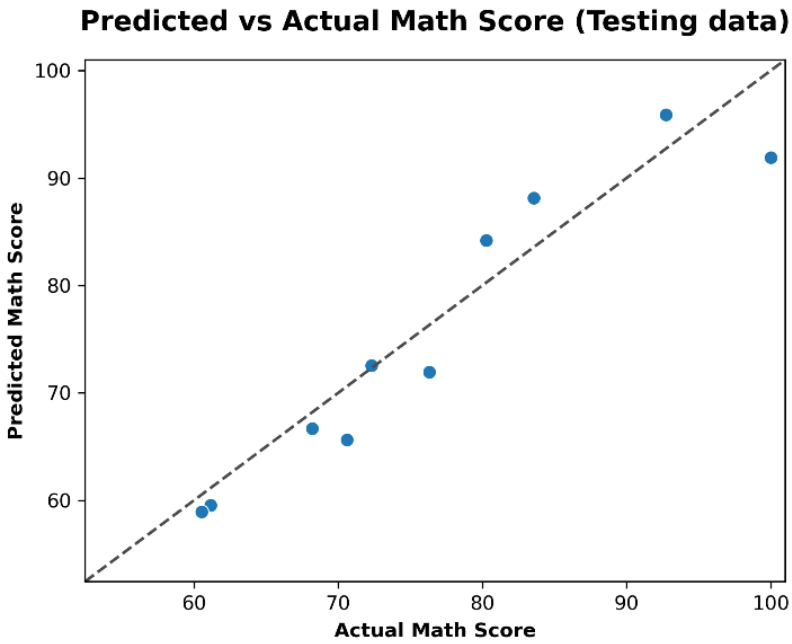


FIGURE 4. Random Forest Regression on Math Score: testing data

Figure 4 shows the performance of the model on the missing observations for math scores. The points follow a reasonably close diagonal trend, although there is more dispersion than in the training data. The model maintains predictive ability across the score spectrum (60-100), with most predictions close to the true values. Significant deviations appear at some points, indicating some predictive error, which is typical for the test data. Overall, the model shows adequate generalization without severe over fitting, although the accuracy decreases compared to the training performance.

TABLE 2. Performance Metrics of Random Forest Regression on Math Score (Training Data and Testing Data)

Parameter	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	Max Error	MSLE	Med AE
Math Score	Train	RFR	Random Forest Regression	0.97983	0.97983	3.61962	1.90253	1.48207	4.62735	0.00071	1.13815
	Test	RFR	Random Forest Regression	0.89114	0.89801	16.24864	4.03096	3.39738	8.07290	0.00234	3.55485

Based on the pattern established from the English and Science score models, the Random Forest regression for the Math score in Table 2 is expected to show a significantly higher fit. The training data will exhibit very strong performance, with a high R^2 value (possibly above 0.95) and low error measures such as RMSE and MAE. In contrast, the performance on the test data will be significantly weaker, with a significantly lower R^2 value and very high errors. This consistent trend across all subjects confirms that while the Random Forest model is powerful at capturing complex patterns in the training set, it struggles to effectively generalize its predictions to unseen test data.

Random Forest Regression (Science Score)

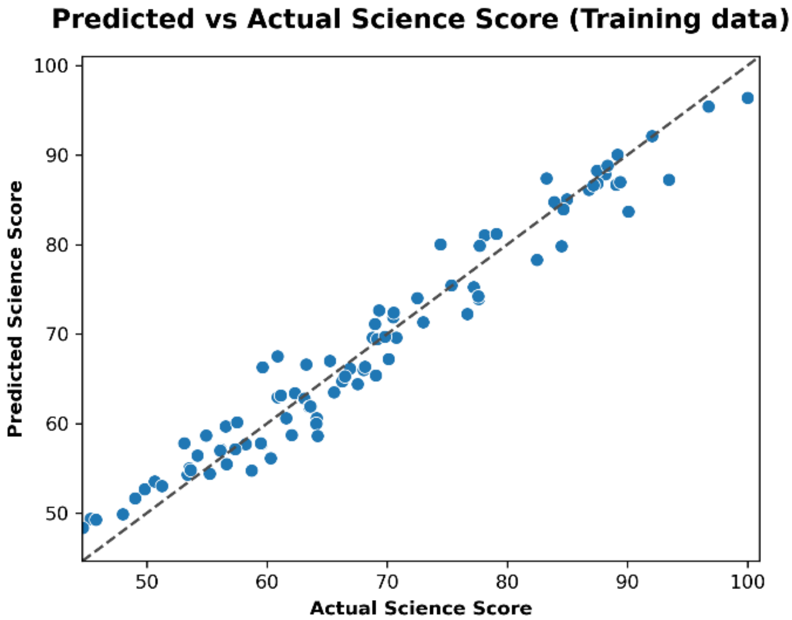


FIGURE 5. Random Forest Regression on Science Score: training data

Figure 5 this scatterplot demonstrates the robust predictive performance of a random forest regression model on science scores. The points are tightly clustered along the diagonal reference line, indicating that the predicted values closely match the actual scores over the entire range (approximately 50–100). The model shows consistent accuracy with minimal systematic bias from low to high scores, indicating that it effectively captures the underlying patterns in the training data and generalizes well across different performance levels.

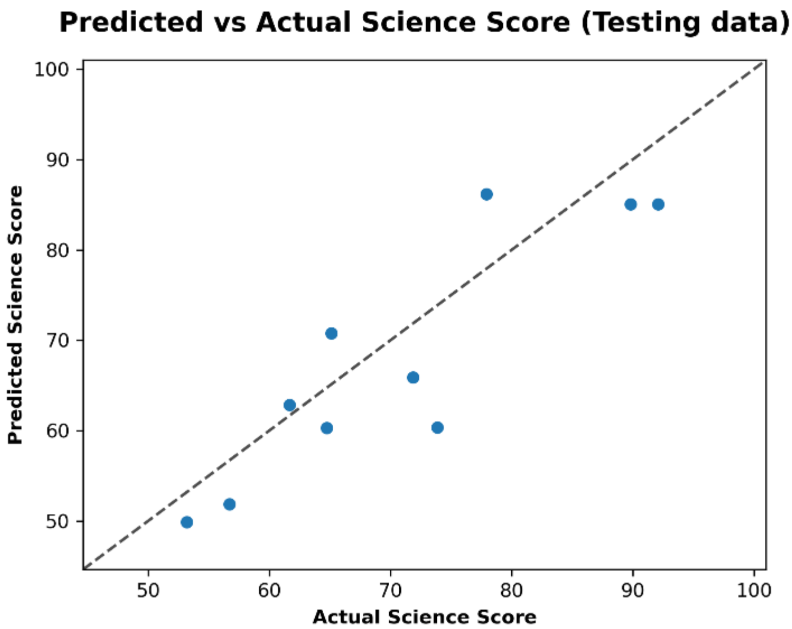


FIGURE 6. Random Forest Regression on Science Score: testing data

Figure 6 this test data plot demonstrates the generalization ability of the random forest model to unobserved data. Although the points generally follow a diagonal trend, there is a significantly higher scatter compared to the training data, with many predictions deviating from the true scores. The model maintains reasonable accuracy over the score range (50-95), although some over predictions and under predictions occur. This increased variability in the test data is expected and indicates that the model generalizes well enough without severe over fitting, although prediction accuracy is reduced.

TABLE 3. Performance Metrics of Random Forest Regression on Science Score (Training Data and Testing Data)

Parameter	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	Max Error	MSLE	Med AE
Science Score	Train	RFR	Random Forest Regression	0.95548	0.95548	8.06948	2.84068	2.33114	6.66900	0.00188	1.99063
	Test	RFR	Random Forest Regression	0.71018	0.76401	44.50243	6.67101	5.88803	13.50480	0.00894	5.25618

The performance metrics for the random forest regression model on the science score reveal a pronounced phenomenon of over fitting, as shown in Table 3. This model shows excellent performance on the training data, with a nearly perfect R^2 of 0.955 and low error values (RMSE: 2.841, MAE: 2.331). However, its performance drops significantly when applied to the unseen test data. The test R^2 drops significantly to 0.710, and all error metrics deteriorate significantly, with the RMSE doubling to 6.671 and the MAE rising to 5.888. This large performance gap indicates that the model memorizes the training data patterns too closely, compromising its ability to generalize to new data.

Random Forest Regression (English Score)

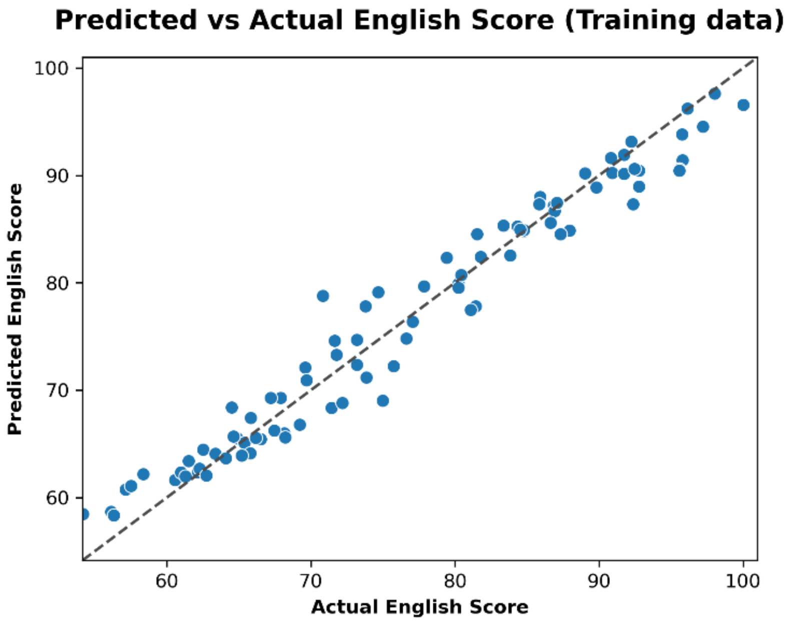


FIGURE 7. Random Forest Regression on English Score: training data

Figure 7 shows that the Random Forest model exhibits strong predictive performance for English scores on the training data, with points clustering closely around the diagonal line over the entire range (55–100). The model shows consistent accuracy with minimal deviation from low to high scores. Some moderate scatter appears in the range 70–85, but the overall fit is tight, indicating that the model effectively captures the patterns in the data and does not show significant systematic bias across different performance levels.

TABLE 4. Performance Metrics of Random Forest Regression on English Score (Training Data and Testing Data)

Parameter	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
English Score	Train	RFR	Random Forest Regression	0.95820	0.95825	6.23557	2.49711	1.96593	7.97855	0.00109	1.57845
	Test	RFR	Random Forest Regression	0.79274	0.79455	30.55703	5.52784	4.42905	10.62100	0.00615	4.13175

Based on the performance metrics in Table 4, the Random Forest Regression model predicting English score shows a significant difference in performance between the training and test datasets, indicating over fitting. On the training data, the model shows exceptional performance, with an R^2 of 0.958, meaning it explains more than 95% of the variance. This is further supported by very low error metrics, such as an RMSE of 2.497 and an MAE of 1.966. However, its performance decreases on the unobserved test data. The test R^2 drops to 0.793, and the error metrics increase significantly, with the RMSE doubling to 5.528 and the MAE rising to 4.429. This significant gap confirms that although the model learns the training patterns well, its ability to generalize to new data is significantly weaker.

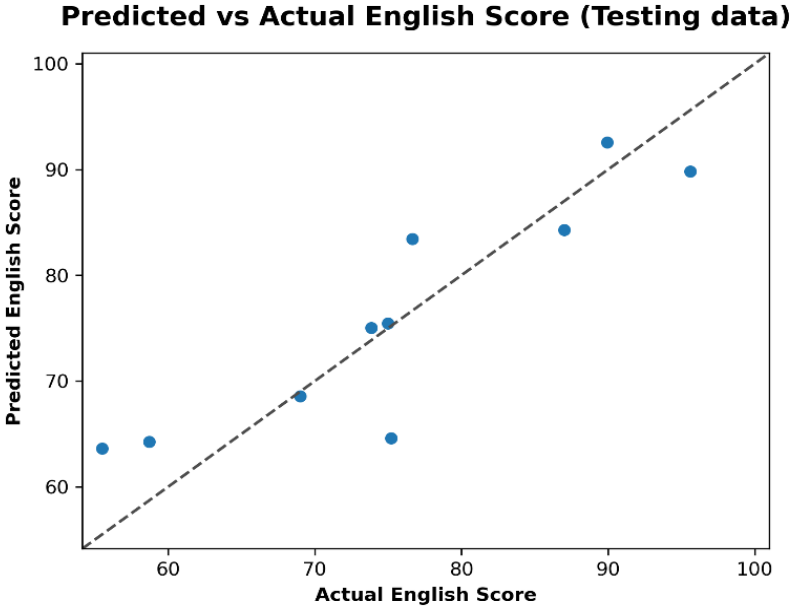


FIGURE 8. Random Forest Regression on English Score: testing data

Figure 8 shows the model's generalization to missing observations for the English test scores. The points are generally aligned along the diagonal, although with increased scatter compared to the training data. The model maintains reasonable predictive accuracy across the scores (55–95), with most predictions approximating the true values. Some significant deviations occur, particularly in the lower and upper bounds, indicating the general predictive variability of the test data. Although accuracy naturally decreases with new data, the model demonstrates acceptable generalization without serious over fitting.

4. CONCLUSION

Demonstrates a significant relationship between artificial intelligence use in educational settings and academic performance, revealing that increased AI-integrated study time is positively associated with student achievement across math, science, and English. The research establishes that students who use AI tools for long periods of time (9+ hours) consistently perform better than those with minimal engagement, highlighting the potential of AI as a powerful educational improvement tool. Using random forest regression analysis, the study successfully predicted academic outcomes with high accuracy on training data ($R^2 > 0.95$), although generalizing to test data showed expected reductions in performance, confirming the validity of the model without severe over fitting. The findings underscore both the promise and challenges of AI integration in schools. While AI technologies offer unprecedented opportunities for personalized learning, automated administration, and data-driven decision-making, significant barriers remain. Infrastructure limitations, inadequate professional training, data privacy concerns, and resistance to innovation continue to hinder widespread adoption. The modest variation in student performance (~13 standard deviations for scores) suggests that AI alone cannot address all educational disparities without strategic implementation frameworks. Going forward, educational institutions should prioritize targeted investments in AI infrastructure, comprehensive teacher training programs, and robust data protection measures. The success of AI in education depends not only on embracing the technology, but also on maintaining human-centred teaching approaches that balance methodological efficiency with empathetic teaching practices. As AI continues to reshape educational landscapes, stakeholders must collaborate to create inclusive, sustainable learning environments that utilize technology while preserving educational integrity. This research provides valuable insights into the transformative potential of AI and provides a foundation for future studies that explore optimal integration strategies that maximize student outcomes while addressing ethical and practical implementation challenges.

REFERENCES

[1]. Mappisabbi, A. M. F., Amirullah, N., & Batti, S. (2024). Leveraging Data and Artificial Intelligence for Innovative Educational Management Practices. Deleted Journal, 2(2), 182–192. https://doi.org/10.54066/ijmre-itb.v2i2.1788

[2]. Alsbou, M., & Alsaraireh, R. A. I. (2024). Data-Driven Decision-Making in Education: Leveraging AI for School Improvement. 1, 1–6. https://doi.org/10.1109/ickecs61492.2024.10616616

[3]. Kodir, A. (2025). Peran Artificial Intelligence (AI) dalam Meningkatkan Layanan Pendidikan di SMP/MTs. Manajemen Kreatif Jurnal, 3(1), 95–104. https://doi.org/10.55606/makreju.v3i1.3622

[4]. Cabrera, E., Batista, E., Palau, R., Unciti, O., Ferré, M., & Martínez-Ballesté, A. (2024). The Use of Artificial Intelligence Techniques in Smart Classrooms is in its Infancy. IEEE Access, 12, 125179–125193. https://doi.org/10.1109/access.2024.3454372

[5]. Alsbou, Majida Khalaf Khaleel, and Ra'Ed Abdalhafed Ibrahim Alsaraireh. "Data-driven decision-making in education: Leveraging AI for school improvement." In 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), vol. 1, pp. 1-6. IEEE, 2024.

[6]. Perrotta, Carlo, and Neil Selwyn. "Deep learning goes to school: Toward a relational understanding of AI in education." Learning, media and technology 45, no. 3 (2020): 251-269.

[7]. AlHalabi, Shaza. "The Role of Leadership in Utilizing AI for School Improvement."

[8]. Khosravi, Hassan, Shazia Sadiq, and Sihem Amer-Yahia. "Data management of AI-powered education technologies: Challenges and opportunities." Learning Letters 1 (2023): 2-2.

[9]. Huang, Lan. "Ethics of artificial intelligence in education: Student privacy and data protection." Science Insights Education Frontiers 16, no. 2 (2023): 2577-2587.

[10]. Vidhya Prasanth, Chitra Periyasamy, M. Ramachandran, Libiya Saravanan, "Biomedical Applications of Polymer-Composite Materials using Fuzzy ARAS Method" Journal on Materials and its Characterization, 4(4), 2025, 4-14.

[11]. Zhu, Tiffany, Kexun Zhang, and William Yang Wang. "Embracing AI in education: Understanding the surge in large language model use by secondary students." arXiv preprint arXiv: 2411.18708 (2024).

[12]. Fullan, Michael, Cecilia Azorín, Alma Harris, and Michelle Jones. "Artificial intelligence and school leadership: challenges, opportunities and implications." School Leadership & Management 44, no. 4 (2024): 339-346.

[13]. IDU, Anne Debekeme. "Utilization and Challenges of Artificial Intelligence in Secondary School Administration in Bayelsa State, Nigeria." FUO-Journal of Educational Research 5, no. 1 (2025): 127-134.

[14]. Cruz-Jesus, Frederico, Mauro Castelli, Tiago Oliveira, Ricardo Mendes, Catarina Nunes, Mafalda Sa-Velho, and Ana Rosa-Louro. "Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country." Heliyon 6, no. 6 (2020).

[15]. Vidhya Prasanth, Mythili Senthil, M. Ramachandran, Soniya Sriram, "Optimization of Electrocoagulation Process Parameters for Enhanced Removal of Heavy Metals from Wastewater: A Comprehensive Review and Experimental Study" REST Journal on Advances in Mechanical Engineering, 5(1), 2026, 1–11.

[16]. Süße, Thomas, and Maria Kobert. "Generative AI at School-Insights from a study about German students' self-reported usage, the role of students' action-guiding characteristics, perceived learning success and the consideration of contextual factors." Zenodo (2023).

[17]. Nguyen, Nathan D. "Exploring the role of AI in education." London Journal of Social Sciences 6 (2023): 84-95.

[18]. Williamson, Ben, Felicitas Macgilchrist, and John Potter. "Re-examining AI, automation and datafication in education." Learning, media and technology 48, no. 1 (2023): 1-5.

[19]. Anam, Khaerul, Muhamad Sadli, and Hadi Wijaya. "Analysis of artificial intelligence (AI) utilization for improving motor skills learning outcomes among elementary school teacher education (PGSD) students." DIAJAR: Jurnal Pendidikan dan Pembelajaran 3, no. 2 (2024): 202-209.

[20]. Gocen, Ahmet, and Fatih Aydemir. "Artificial intelligence in education and schools." Research on Education and Media 12, no. 1 (2020): 13-21.

[21]. Gocen, Ahmet, and Fatih Aydemir. "Artificial intelligence in education and schools." Research on Education and Media 12, no. 1 (2020): 13-21.

[22]. Rahmatika, Widya, M. Mursalin, Eri Saputra, and Mutia Fonna. "The Influence of the Use of AI in Mathematics Learning at the High School Level: A Systematics Literature Review." Electronic Journal of Education, Social Economics and Technology 6, no. 2 (2025): 1068.