March 2026, Vol:4, Issue:1

International Journal of New Findings in Engineering, Science and Technology

journal homepage: *https://ijonfest.gedik.edu.tr/*

Received:03.02.2026 Accepted:17.03.2026 Published:26.03.2026

Predicting Student Academic Performance Using Machine Learning: A Case Study On Educational Data From Bangladesh Towards Sustainable Education (SDG 4)

Tuka Albonny[a], Izzet Parug Duru[b*]

[a]*Istanbul Gedik University, Statistics and Data Science program*, *Istanbul, Türkiye, tukayasser135@gmail.com*
[b]*Istanbul Gedik University, Medical Imaging Techniques, Istanbul, Türkiye, parug.duru@gedik.edu.tr (*Corresponding Author)*

Abstract

Educational sustainability is a global priority. Because of the rapid digital transformation that we are seeing, it is necessary to integrate modern technologies to achieve inclusive, equitable, and high-quality education. Despite this, there is still a limitation in applying Machine Learning (ML), as one of the modern technologies, to educational data within institutions. This research addresses this gap by providing practical evidence that predicting academic performance via ML is both achievable and highly effective. Two main objectives were generated from the main goal. The first one is to compare four ML models, while the second is to determine the most important factors affecting academic performance. To achieve the previous goals, the study used data from more than 8,000 secondary school students from Bangladesh. Based on the given dataset, students' academic performance was represented by the average grades of five common courses, where the study tried to predict it without relying on the existence of previous grades. The study depends on the contextual, behavioral, and other variables that are available in that dataset. Three of the models used were shallow: Multiple Linear Regression (MLR), Decision Tree (DT), Random Forest (RF), while one was deep: Feedforward Deep Neural Network (DNN). The whole process was implemented in the R environment, and the results revealed some interesting points. The shallow RF model was more accurate than the deep one with a small margin ($R^2 = 93.3$ vs. $R^2 = 93.0$). The most important factor affecting student performance is the student group (arts, commerce, or science), which dominated the predictive power of the models. Study time and attendance can also not be ignored as important behavioral modifiable factors. In addition to the results presented, this study contributes to educational sustainability by developing two Early Warning Systems (EWS): a simplified system for quick group screening and a full EWS for individual predictions. The results and the contributions of this study will help to find the most at-risk subgroups, particularly in the arts track, and allocate resources to provide proactive support based on modifiable behaviors or other factors. Ultimately, this research provides a scalable framework to enhance equity and efficiency in education. It is aligned with UN Sustainable Development Goal 4 (SDG 4): Quality Education.

Keywords: Educational sustainability; Machine learning; Early warning systems; SDG 4; Random forest.

1. INTRODUCTION

Ensuring inclusive and equitable quality education is a cornerstone of global sustainability, as outlined in the United Nations Sustainable Development Goal 4 (SDG 4). In the digital age, this sustainability can be achieved by leveraging the intersection of different fields. In this context, Education and Data Science come up as two areas that can be linked. Being a teacher on one hand and a student in Statistics and Data Science on the other has generated strong motivation to combine these two areas, where analytical tools provided by Data Science can be used in an educational context to improve the quality of education in the long term. One of the analytical tools used in Data Science is Machine Learning. Machine Learning is a set of statistical algorithms that can learn from data and then generate deep insights from it. The main tasks of Machine Learning are to predict future outcomes and classify objects into specific categories [1]. Machine Learning capabilities are employed widely in the field of education. The algorithms of ML are used to personalize educational content and analyze student behavior [2]. Machine Learning applications have been extended to include educational recommendation systems, plagiarism detection, predictive analytics, adaptive learning platforms, and intelligent tutoring systems [3]. ML algorithms are also used to assess interactions in virtual

https://doi.org/10.61150/ijonfest.2026040105

classrooms and offer immediate support to students [4]. Despite the significant potential of Machine Learning (ML) in education, a clear gap persists between technical feasibility and practical implementation. This study addresses this gap by demonstrating the effectiveness of ML in predicting student academic performance, providing a scalable and practical framework that educational institutions can readily apply to their own data.

Based on this, the study aims to answer two questions: The first one is to what extent machine learning models (MLR, DT, RF, and Feedforward DNN) accurately predict academic performance in foundational common subjects when relying solely on contextual and behavioral data (i.e., no prior grades)? This includes determining which model achieves the highest accuracy and whether the deep learning model outperforms the others. The second question is, what are the most important factors affecting the performance of students as identified by the models?

This research is significant; from a theoretical perspective, this research serves as an additional study in existing literature. Specifically, it represents additional research on predicting academic performance using Machine Learning. This research can be used by other researchers to learn from or build on it. Practically, it provides a clear methodology for teachers. The steps provided in this research guide them on how to use ML algorithms on educational data. On a professional level, conducting this study equips educators with advanced competencies in data literacy, algorithmic thinking, and the use of analytical tools like R. These practical skills are in high demand and relatively uncommon in the modern educational context. Furthermore, this research opens new horizons for exploring further applications of ML in the education sector.

2. LITERATURE REVIEW

In this section, we begin by reviewing some systematic reviews and foundational studies, and then we cover some of the recent related studies. At the end of this section, the research gap will be identified.

2.1 Foundational and Systematic Reviews

Here, we present two systematic reviews. Both showed that the prediction of student performance is a dynamic and active field, but still, there are many gaps. In [5], the researchers reviewed 87 studies published between 2009 and 2021. Many gaps were identified in the reviewed studies: using a limited data size, focusing on binary classification tasks, and neglecting to handle imbalanced data. The researchers indicated that the studies used a wide range of algorithms; most of them applied traditional ML algorithms like DT, Naive Bayes, and Support Vector Machines (SVMs), with only a few applying deep learning algorithms. Regarding the variables used in the studies reviewed, they used static data (like university records) or dynamic data (from online learning platforms).

The study in [6] is newer compared to the earlier one. Researchers examined studies published from 2011 to 2023. They agreed with [5] about using various algorithms, from simple to complex, in the examined studies. They also agreed with the previous review on the limited data size; they stated that studies used datasets with sample sizes from 89 to 2400. They confirmed that there is no one best algorithm. The best model differs as studies differ. In terms of variable importance, the Grade Point Average (GPA) was used the most, and it was the most impactful variable, followed by past grades and demographic/socioeconomic factors.

When literature reviews are conducted, foundational studies should not be ignored. So, alongside the systematic reviews, we highlighted one important classical study: [7]. In this study researcher got a real-world dataset from two Portuguese secondary schools. They wanted to predict performance in math and Portuguese. Several algorithms —DT, RF, Artificial Neural Networks (ANN), and SVM—were compared across three distinct tasks: binary classification, five-level classification, and regression. The findings indicated that the Decision Tree performed best in classification tasks, while Random Forest was best for regression. This study was significant not only for showing that tree-based models (DT and RF) were superior at the time but also for being one of the first to show how important non-academic variables are. It found that the most accurate predictor was previous student performance (previous grades); however, in the absence of these grades, alternative factors such as past failures and school absences became significantly important.

2.2 Analysis of Applied Studies

The field of Educational Data Mining (EDM) and Learning Analytics (LA) has witnessed a significant qualitative evolution in recent years. While earlier applied studies primarily established the efficacy of using static academic records, historical grades, and GPA as the main predictors (e.g., [8, 9]), some more recent research has shifted toward prioritizing dynamic behavioral indicators (e.g., [11, 12, 15]). These include engagement metrics such as resource interaction, attendance, and online learning activities, which are viewed as more responsive predictors for creating effective Early Warning Systems (EWS). Alongside this shift in variables, there is a clear modern trend in the literature toward adopting Deep Learning (DL) architectures, where these models proved to be the most accurate and effective in handling large-scale and complex educational datasets (e.g., [12]).

Table 1 presents a selected sample of applied studies conducted between 2020 and 2025. This collection highlights the diverse methodologies and predictive approaches used in the field.

Table 1. Summary of applied studies.

Study	Sample Size	Key Predictors	Target Variable	Best model	Accuracy/ Result
[8]	250	continuous assessment grades (include Quiz1, Quiz2, Midterm, Project, and Lab).	final grade (A–F).	Linear Discriminant Analysis (LDA)	81% (Reached 90.7% on independent test)
[9]	499	demographic and academic variables (e.g., activity points and exam points).	actual grade (6 categories)/ pass/fail (2 categories).	Logistic Regression in both classification tasks.	88.8% (Binary) / 68.7% (multi-class)
[10]	2161	There were 41 variables. Attendance and Student activities were the most important.	Final grade.	DT (Classification) / AdaBoost (Regression)	98.46% accuracy / 0.915 R^2
[11]	480	demographic, academic, and behavioral (e.g., raised hands, visited resources, and additional metrics).	student's performance level (low, medium, or high).	Voting (ANN + BayesNet)	81.6% accuracy.
[12]	32593	demographic, on-platform activity, and prior educational history.	Binary (e.g., Pass/Fail, Withdrawn/Pass).	Deep ANN	94.7% (Withdrawal) / 84.4% (At-risk)
[13]	4266	1st-year course grades, failures.	Data Structures grade (2nd Year, Pass/Fail).	DNN	89% (Superior to DT, RF, and LR)
[14]	Not specified	test scores, course grades (from early-level courses), attendance records, and continuous evaluation scores.	pass or fail	DNN	91.0% (Superior to SVM and DT)
[15]	300	Behavioral indicators: (Repeated views of resources **RV-N**, Resource utilization efficiency **RU-E**, Resource density utilization **RD-U**).	Binary classification: Student performance (Excellent [score > 90] / Not excellent)	Logistic regression model with Taylor expansion	93.3% accuracy on the testing set

2.3 Research Gap

The preceding literature review shows that the field of academic performance prediction, while a dynamic area of research, is characterized by conflicting results and instability. Most research has focused on comparing traditional shallow models, and their conclusions remain inconclusive, as the "best" algorithm differs by study. Additionally, after the studies discussed in the previous two sections, we can clearly note that there is a trend toward the use of deep models. However, this area still needs more comparison between shallow and deep models to see if deep learning will always be the best or not.

We should not ignore the use of a limited dataset size, as it represents a gap. Studies like [6] pointed out this problem. When the data size is very small, the reliability of the results decreases. Furthermore, it seems that the studies related to academic performance prediction in areas like Bangladesh are relatively rare. The rarity leads to limiting our understanding of the effectiveness of the models across different cultural and economic environments.

In addition to that, it was also noted that most studies looked at the problem as a classification task (pass/fail), while only a few studies looked at the regression task (predicting the numerical grade). Furthermore, many studies focused on predictive accuracy without providing an analysis of the most important variables, therefore reducing the interpretability and practical utility of the results to educators.

Notably, most existing literature predicts performance either in individual specialized courses or as a cumulative GPA. There is a distinct lack of studies that focus specifically on predicting the average performance in foundational common subjects that serve as the universal core across different academic streams. Finally, many prior studies relied on "prior academic grades" as a strong predictor (as noted by [6,13]), and prediction was not attempted in the absence of these grades. This reliance increases accuracy, but at the same time, it makes the models less useful as "Day 1" early warning systems before any grades are recorded.

To fill the gaps that have been noted, this study will compare shallow with deep models (MLR, RF, and DT vs Feedforward DNN), focusing on the regression task for the average score in five foundational common subjects. It will address the context and size gap by using a dataset from Bangladesh containing over 8,500 records. It will predict academic performance relying only on behavioral and contextual variables, with the dual aim of understanding the most influential factors and applying a model with practical utility as a preliminary screening tool.

3. METHODOLOGY

This section explains the entire workflow that was followed to achieve the research goals.

3.1 Data Source and Context

The dataset was obtained from the Kaggle platform. It contains information on secondary school students in Bangladesh. In the Bangladeshi system secondary stage begins at grade 9, where students choose one of three tracks: Science, Commerce, and Arts/Humanities [16]. There are common compulsory subjects for secondary students such as English, Mathematics, Social Science, and Information and Communication Technology [16,17].

To ensure a balanced and representative sample, a hybrid data collection strategy was employed. Eighty percent of the data was collected through a custom-built website form, allowing for efficient, large-scale collection from students residing in diverse regions of Bangladesh. The remaining 20% was gathered through direct physical interactions at three different schools, utilizing printed questionnaires and in-person interviews to ensure high-fidelity responses. All physically collected data was subsequently manually digitized to maintain rigorous consistency with the digital entries [18].

The dataset, as it was extracted from its source, contains 8,612 records and 24 features, covering various types of variables, including demographic, behavioural, and academic variables. A new variable, the 25th variable, was added: the Average Score, after reading the dataset in the R environment. We calculated this variable based on the grades of five core subjects available in the dataset, namely: English, Math, Science, Social Science, and Art Culture.

3.2 Exploratory Data Analysis (EDA) and Preprocessing

The EDA revealed some problems that should be fixed. The dataset contained age values of 10 and 24, both lie outside the boundaries of the target population (14-19). In addition, the minimum value in the family size variable was 0. The categories of the location variable were inconsistent (e.g., urban, Urban, rural, city, City), and an empty cell in this variable was found. The problem of inconsistency also appeared in the variables mother_education and father_education. Moreover, there were 315 fully duplicated rows. No explicit missing values (NAs) were found in the dataset.

3.3 Data Preprocessing and Preparation

To address the problems that emerged in the EDA and ensure model reliability, a multi-step preprocessing pipeline was implemented. Initially, we excluded non-predictive variables, including id and full_name, along with the five individual subject scores used to calculate the target variable. Furthermore, 315 fully duplicated rows were removed to ensure data integrity. Data cleaning involved filtering out outliers in the age variable (values 10 and 24) and removing records with an invalid family_size of 0. Inconsistent categorical entries in location and parental education variables were standardized to ensure uniformity. Empty cells were treated as NA, and the na.omit() function was applied to remove any remaining incomplete records.

Following the cleaning process, categorical variables were converted into factors for the regression and tree-based models. The final dataset (8287 × 18) was partitioned into an 80% training set and a 20% test set. For the Feedforward Deep Neural Network (DNN), specific preparation steps were conducted, including One-Hot Encoding and Min-Max normalization to scale all predictors within a [0, 1] range, and converting the datasets into matrix format for use with Keras.

3.4 Visual Data Analysis

In this phase, two plots were generated to explore the relationships between the variables. Figure 1 and Figure 2 show these relationships.

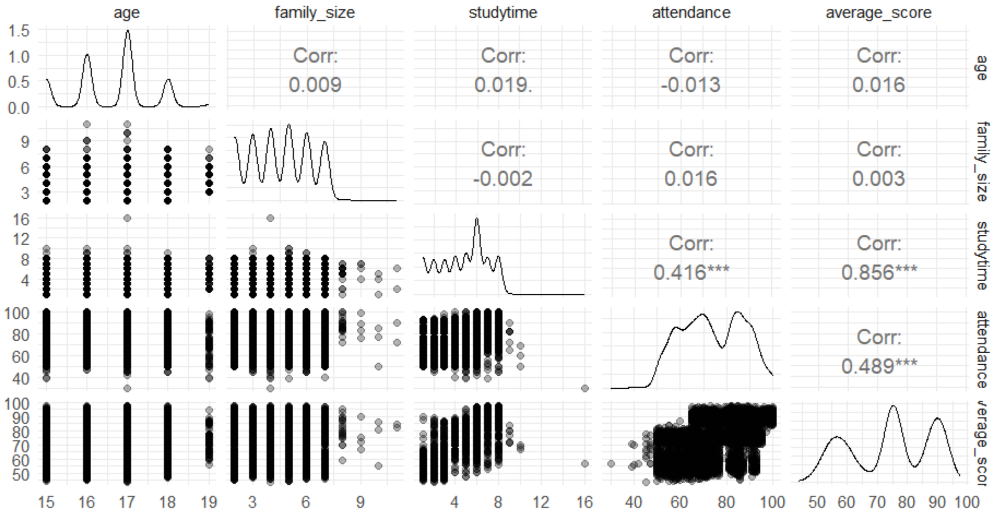

Figure 1. *Scatterplot matrix and correlations.*

As shown in Figure 1, Study time has the highest correlation coefficient with the target variable (0.856), followed by attendance (0.489). This high level of association provides an initial indication that both variables are likely to be highly predictive of core subject performance. Additionally, there was a moderate positive correlation of 0.416 between study time and attendance. This correlation suggests that there is a behavioral link between the two.

Boxplot analysis (Figure 2) revealed that Science track students achieved higher scores compared to other groups, which means that the academic track could be a strong predictor of academic performance in these five common subjects. In contrast, variables such as mother_job, school_type, and internet_access seem to have weaker links.

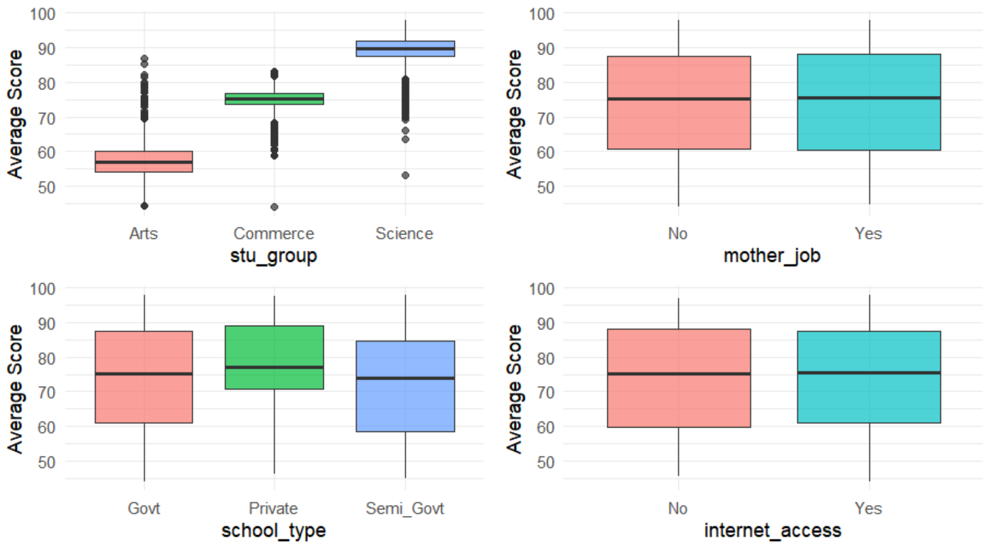

Figure 2. *Boxplot analysis of average score across some categorical predictors.*

3.5 Predictive Models

Four supervised ML models were constructed and compared.

1. Multiple Linear Regression (MLR): used as a baseline model, its assumptions were verified.
2. Decision Tree (DT): provides easy and clear rules to understand.
3. Random Forest (RF): combines many decision trees to improve accuracy.
4. Feedforward Deep Neural Network (DNN): used to compare with the shallow models.

In DT, RF, and Feedforward DNN models, we ensured meticulous tuning of hyperparameters to achieve the highest possible predictive accuracy.

Specifically, the tuning process for each model was conducted as follows:

- **Decision Tree (DT):** The model was optimized via 10-fold cross-validation to select the best Complexity Parameter (cp) from 20 candidate values.
- **Random Forest (RF):** A manual grid search was performed to tune mtry (tested values: 5, 7, 9, 11, 13), min.node.size (1, 3, 5), and sample.fraction (0.632, 0.8). The optimal configuration reached was mtry = 7, min.node.size = 3, replace = TRUE, and sample.fraction = 0.8.
- **Feedforward DNN:** A random search of 50 combinations was implemented to optimize the number of units per hidden layer (64, 128, 32), dropout rates (0.4, 0.3, 0.0), and learning rates. To enhance stability and prevent overfitting, Batch Normalization and Early Stopping (with a patience of 10 epochs) were utilized.

To ensure the robustness and generalizability of the results, a dual validation strategy was adopted. First, Internal Validation was performed during the training phase using specific techniques for each model: 10-fold cross-validation for the Decision Tree, Out-of-Bag (OOB) error estimation for the Random Forest, and a 20% validation split with early stopping for the DNN. Second, External Validation was conducted by evaluating all final tuned models on a strictly independent test set (20% of the data) that was not seen by the models during training or hyperparameter tuning.

3.6 Performance Evaluation Metrics

We use three metrics in the evaluation process. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R-squared):

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

$$MAE = \frac{1}{m}|y_i - \hat{y}_i| \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \tag{3}$$

3.7 Practical Scenario Testing and Early Warning System (EWS) Development

To make the best model a practical tool, an early warning system was developed. This model works on the first day of the semester to identify students at risk. It was constructed by excluding variables collected later in the semester. The early warning model was later simplified to include the most critical variables.

4. RESULTS

4.1 Multiple Linear Regression (MLR) Results

The MLR model achieved a high R^2 on test data, reaching approximately 92%. However, when the linearity assumption was tested, it was found to be violated, which means that the relationships between predictors and the target variable were not linear. Additionally, there was a violation of the normality and homoscedasticity assumptions. These violations make the MLR model statistically unreliable. These findings justify the move toward non-linear models like Random Forest and DNN.

4.2 Decision Tree Results

Figure 3 illustrates the structure of the decision tree.

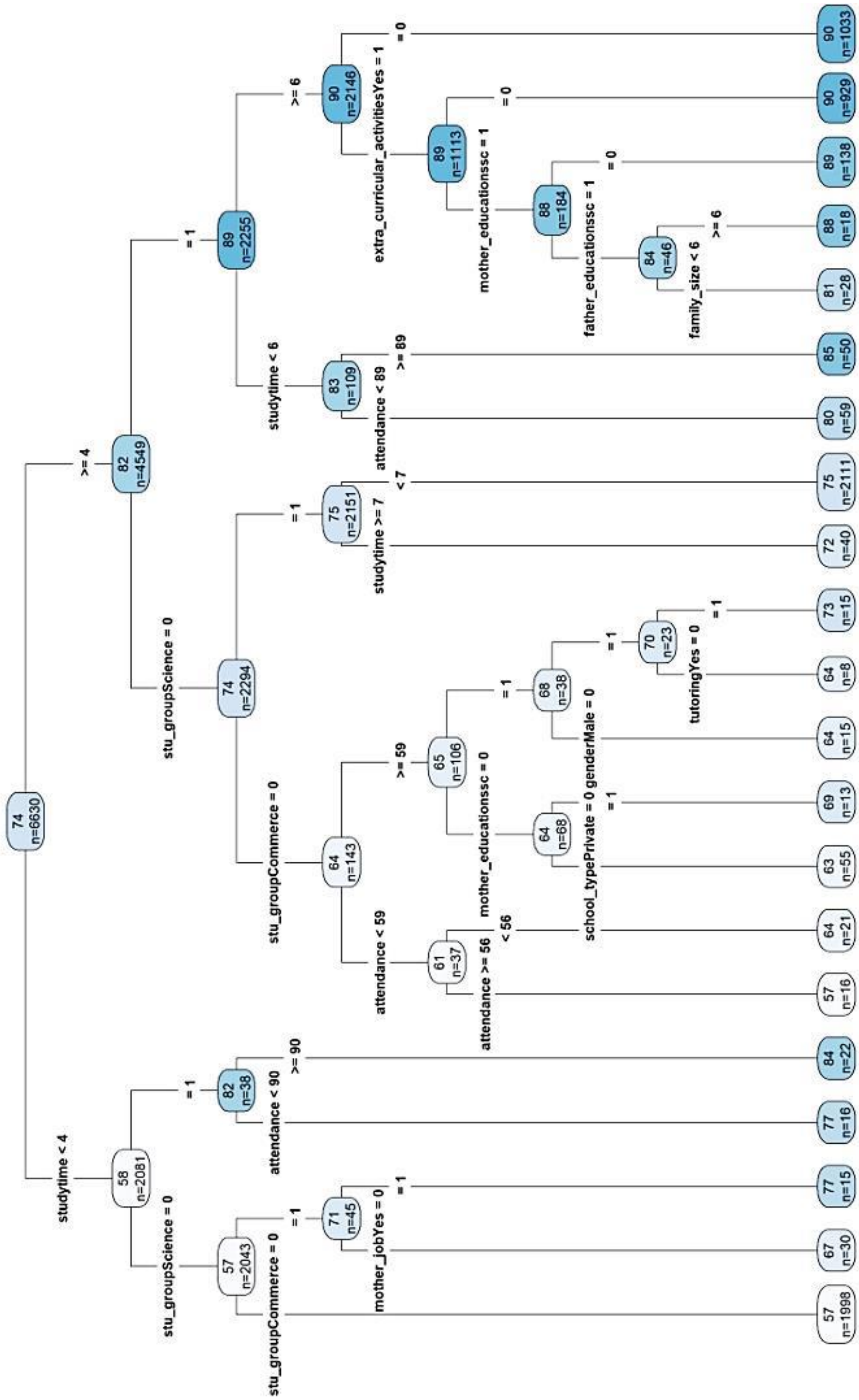

Figure 3. The Decision Tree Structure.

The chart shows that Arts students had the lowest average in the core subjects, with scores ranging from 57 to 73. For Commerce students, the average scores in the core subjects were higher, ranging from 67 to 77. When study hours were between four and seven, the predicted score was 75, which was the expected score for most Commerce students. The highest grades were all achieved by science students. For instance, the maximum grade of 90 appeared in two distinct paths of the decision tree. The first path included students who studied for six hours or more and did not participate in extracurricular activities. The second path included students who studied for six hours or more, participated in extracurricular activities, and whose mothers did not hold a Secondary School Certificate (SSC).

As for the importance of the variables in the tree model, they are shown from most important to least important in Figure 4.

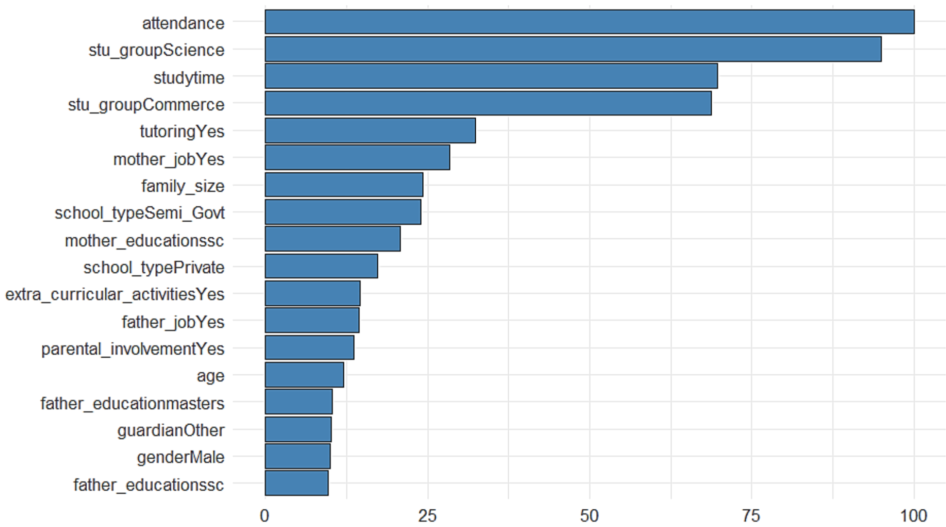

Figure 4. *Variable importance of the Decision Tree Model.*

By summing the importance scores of stu_groupScience and stu_groupCommerce, the overall importance of the stu_group variable exceeds that of attendance, making it the most significant predictor. Consequently, study time ranks third. Regarding the remaining variables, some showed moderate importance, such as tutoringYes and mother_jobYes, while demographic factors (e.g., age and gender) contributed very little to the model's predictive power.

4.3 Random Forest Results

To assess the factors relied upon by the Random Forest (RF) model, the predictor importance was analyzed, as shown in Figure 5.

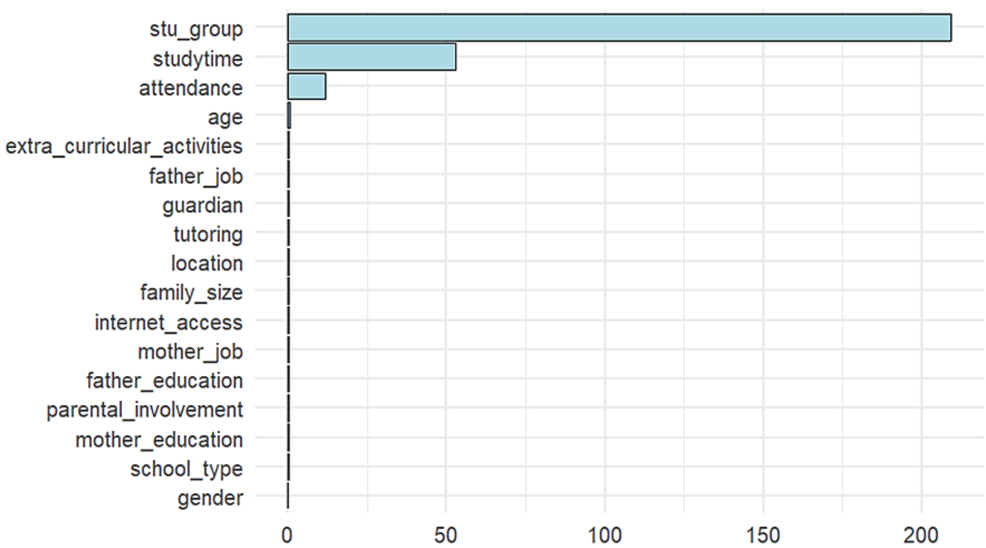

Figure 5. Variable importance of Random Forest.

Figure 5 clearly shows the dominance of the stu_group (Student Group) variable as the most important predictor in the model by a significant margin. In second place is studytime, followed by attendance in third. A distinct "cliff," or sharp drop-off in importance, is visible after these top three variables. The contribution of the remaining predictors (such as age and extra_curricular_activities) was minimal and approached zero.

4.4 Feedforward DNN Results

Figure 6 illustrates the variables the model found most impactful in making its predictions.

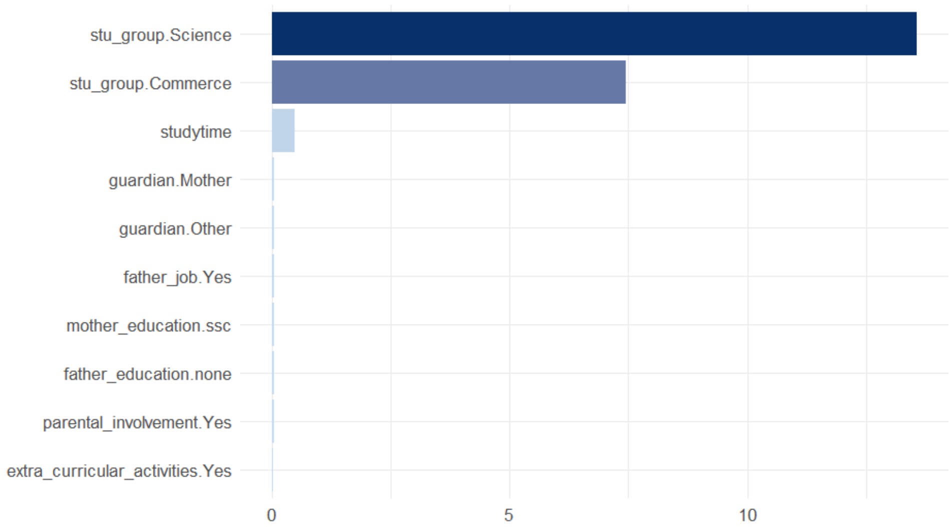

Figure 6. Variable importance for Feedforward DNN.

It shows that the Deep Neural Network model determined that the student's academic group (stu_group) is the only predictor with dominant importance. In contrast, the studytime variable appears with very little importance, and the other variables show almost no importance at all.

4.5 Comparative Performance Metrics of Full Models

Table 2 summarizes the definitive performance metrics for the four models developed.

Table 2. Performance comparison (Test Set).

Model	RMSE	R^2	MAE
Random Forest (RF)	3.555	0.933	2.655
Decision Tree (DT)	3.632	0.930	2.681
Feedforward Deep Neural Network (DNN)	3.635	0.930	2.731
Multiple Linear Regression (MLR)	3.806	0.923	2.816

It is evident from the table that the three non-linear models (DT, RF, and DNN) all achieved excellent performance and were very similar in their results. Notably, all of them outperformed the linear model. However, the Random Forest model is the best. It reached the highest accuracy (Test $R^2 = 0.933$) and the lowest Test RMSE of 3.555.

4.6 Early Warning System (EWS) Scenarios

To make the winning model, which is the RF model, useful in practice, it was used to build the EWS model by removing the attendance, as it is only available at the end of the semester, and re-tuned its hyperparameters. Subsequently, the early warning system was simplified to include only the two strongest predictors of performance, which are student group and study time.

The first EWS achieved R^2 of 0.9317764. Compared to the R^2 of the full Random Forest model, the difference is approximately 0.001, indicating that the early warning model maintained nearly the same predictive power. In this EWS, predictions are even more concentrated on the top two factors: student group, which remains the strongest and first predictor, followed by study time. This model provides a unique prediction for each student.

The R^2 of the simplified EWS model equals 0.9295529. This result proves that the vast majority of the model's predictive power for core subject performance comes from just these two variables. In this system, one predicted value is given for all students who share the same academic track and study hours. For instance, science students who study for seven or eight hours achieve an average of approximately 90. According to this model, arts students who study 1-3 hours are the most at risk of low performance in core subjects. They achieve an average of nearly 57 or less.

5. DISCUSSION

In this section, the answers to the research questions are presented and discussed. The answer to the first question is that machine learning models can predict academic performance in foundational common subjects without prior grades, with a high degree of accuracy, reaching approximately 93%. In terms of best performance, the RF model outperformed the others, including the deep model, with an R^2 of 93.3%. Despite the superiority of the Random Forest model, the Decision Tree can be considered the best explanatory model, as the tree diagram revealed the interactions between primary and secondary variables to produce the final prediction. Regarding the second research question about the most important factors affecting academic performance, it can be said that the student group is the most important factor. The statistical dominance of the 'stu_group' (academic track) variable is deeply rooted in the structural design of the Bangladeshi education system. According to the National Curriculum Framework [16], students are streamed into Science, Business Studies, or Arts /Humanities tracks at the beginning of Grade 9 based on their aptitudes and academic performance. While the system allows for student preference, entry into the Science track is formally regulated by mandatory GPA thresholds achieved in the Junior School Certificate (JSC) examinations at the end of Grade 8 [16, 17]. Consequently, high-achieving students are predominantly concentrated in the Science track, average-achieving students in the Business track, and lower-achieving students in the Humanities/Arts track [17, 19]. This confirms that the 'stu_group' variable functions as a proxy for cumulative prior achievement, explaining its high predictive weight in the models. In fact, the models relied on the arithmetic means of all average scores of each academic track in generating the prediction process. The remaining variables were then utilized by the models to adjust this arithmetic mean and enhance predictive accuracy. After the student group, study time emerged as the most significant modifier and the second most important variable. The attendance variable is the second modifier and the third important factor. The combined contribution of the remaining factors was lower; they slightly improved the predictive accuracy.

When placing our study findings within the context of the existing literature, a notable divergence emerges regarding model superiority. In three of the studies discussed in the literature review section, the deep models were better than the shallow models. These studies were [12], [13], and [14]. The results of our study contradicted the findings in those studies; in our study, the RF model was better, with a difference of 0.3% in test R2. This difference in results between our study and other studies is attributed to the difference in the variables used, including both the predictors and the target variable. Regarding variable importance, there were also some differences. [12] confirmed the role of demographic characteristics, while in our study, the importance of these factors was either minimal or nonexistent. Both [11] and [14] agreed with our study in emphasizing the importance of behavioral factors such as interactions and attendance. Our study added that the importance of study time as a behavioral factor cannot be ignored.

Beyond the academic comparisons, the findings of this study have significant practical implications. First, identifying the study time variable as the most important variable affecting performance has important practical implications. Since students are guided based on their previous performance to choose their academic track, as we discussed previously, providing support for students in core subjects must be at an early stage, before the secondary stage. That is, the intervention should not be limited to the support provided during the student's secondary studies. Second, identifying study time and attendance as significant influence factors is crucial. These two factors are modifiable and improvable, which raises the average performance in common core subjects. Third, the simplified early warning system is a fast diagnostic tool that is easy to understand and use for both teachers and administrators, or for institutions in general. There is indeed a clear ranking of student performance; science students are the best, followed by commerce students, and then arts students. However, this model identifies the weakest subgroup of students within each track. Identifying these groups allows for directing support in the core subjects to the category that needs it most. Fourth, the full EWS model can be used for precise individual diagnosis. Students within a specific track who study for a specific number of hours may have variations in their performance based on other factors (the secondary factors). This model enables us to understand the factors that led to this disparity. Additionally, this model enables teachers and administrators to identify students most at risk of low performance within groups that share the same academic track and study hours. This allows for personalized interventions tailored to each student's specific weaknesses. To bridge the gap between technical modeling and school-level implementation, the provided R code can be deployed via user-friendly platforms such as R Shiny. This would enable administrators to utilize the predict function through a visual dashboard without needing a programming background. By integrating this engine into School Management Systems, institutions can generate immediate at-risk reports at the beginning of each term. Collectively, these practice implications have relevance for classroom results and relate to sustainability in education more generally. Highlighting efficiency, equity, and inclusiveness, these results illustrate that

Early Warning Systems (EWS), which are based on predicting risk in core academic areas, allow for targeted resource allocation to the students who need it and ensure fair support across all learners. Furthermore, the early identification of at-risk students is a proactive action to decrease failure and dropout. This reinforces the continuity and resilience of educational systems, directly advancing progress toward Sustainable Development Goal 4 (SDG 4): ensuring inclusive and equitable quality education for all [20].

While the models demonstrate clear practical value, the findings must also be interpreted with caution due to certain methodological limitations. First, the very high accuracy (~93%) was a result of the presence of the study time variable. This strong variable made the task more akin to a structural classification than to a typical regression. Such a variable may not exist in other databases, and if it does, there may not be clear differences between its categories, which limits the generalizability of the results. Second, we calculated the target variable using a specific subset of five common compulsory subjects: English, Math, Science, Social Studies, and Arts and Culture. A decrease in the value of this variable may be a result of a decline in performance in at least one of the five subjects. Therefore, this variable will not enable us to direct support to each student according to their needs; for example, a student may need additional support in Mathematics more than in the other subjects included in the target variable. Third, certain variables were not included in the dataset used in this study, such as motivation, persistence, family income, and socioeconomic status (SES). Since these variables may affect performance, their absence limits the scope of the analysis. Fourth, regarding the timing of data collection, it is possible that students reported their study habits after becoming aware of their academic results. This introduces the potential for retrospective rationalization, where students might have adjusted their perceived study time to align with their performance. While this is a common limitation in cross-sectional educational datasets, it should be considered when interpreting the strength of the correlation between study time and academic outcomes.

6. CONCLUSION, RECOMMENDATIONS, AND FUTURE WORK

Our study achieved its primary objectives. This study confirmed that academic performance in core foundational subjects can be predicted with a high degree of accuracy ($R^2 \approx 93\%$), even when relying solely on contextual and behavioral variables. Comparisons showed that the Random Forest model was the most accurate, slightly outperforming other models. The analysis of variable importance showed that the student group was the most important factor, followed by study time and attendance.

Drawing on these findings, several practical recommendations can be outlined.

- First, since the academic track was found to be a reflection of prior student quality rather than a direct cause of performance, we recommend that educational authorities in Bangladesh and similar contexts focus intervention efforts on pre-secondary education, particularly for students in the humanities/arts track, whose lower academic achievement compared to other tracks is evident. Strengthening core competencies at earlier stages is more sustainable than focusing solely on remedial programs after specialization.
- In addition to the previous recommendation, schools should systematically monitor attendance and study habits, as they represent "modifiable behaviors" that students and teachers can directly influence to improve outcomes.
- Also, we recommend the implementation of the developed models. The Simplified EWS should be adopted as a rapid screening tool at the beginning of the semester to identify vulnerable subgroups using only track and study time data. The Full EWS should be utilized for individualized diagnosis and in-depth classroom intervention, providing counselors with a "checklist" of secondary factors (e.g., parental occupation and guardian status) to customize support plans.

Ultimately, adopting these data-driven systems serves as a tool to enhance educational sustainability. By ensuring efficiency in resource allocation and reducing dropout rates, institutions move closer to achieving Sustainable Development Goal 4 (SDG4): Quality Education.

Finally, it is suggested that researchers test the generalizability of the models by conducting tests on datasets from different cultural contexts (outside of Bangladesh). We also propose adding factors that were not previously available (such as psychological factors like motivation or economic status) to investigate whether such changes would improve accuracy.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and Code Availability

The complete R source code for this study is publicly available on GitHub to ensure transparency and reproducibility. The repository can be accessed at: https://github.com/tukayasser135-hub/Student-Performance

Authors' Contributions

No	Full Name	ORCID ID	Author's Contribution
1	Tuka Albonny	0009-0004-6556-5667	1,2,3,4
2	Izzet Parug Duru	0000-0002-9227-2497	1
1- Study design 2- Data collection 3- Data analysis and interpretation 4- Manuscript writing			

References

[1] Lewis N D. *Machine learning made easy with R: An intuitive step-by-step blueprint for beginners*. 1st ed. AusCov; 2017

[2] Huang, J., Zhong, Y., Chen, X., 2025. Adaptive and personalized learning in STEM education using high-performance computing and artificial intelligence. *Journal of Supercomputing*, 81(1), 981–1004. https://doi.org/10.1007/s11227-025-07481-7

[3] Romero, C., Ventura, S., 2020. Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. https://doi.org/10.1002/widm.1355

[4] Ersozlu, Z., Taheri, S., Koch, I., 2024. A review of machine learning methods used for educational data. *Education and Information Technologies*, 29, 22125–22145. https://doi.org/10.1007/s10639-024-12704-0

[5] Albreiki, B., Zaki, N., Alashwal, H., 2021. A systematic literature review of students' performance prediction using machine learning techniques. *Education Sciences*, 11(9), 552. https://doi.org/10.3390/educsci11090552

[6] Ouhaddou, C., Retbi, A., Bennani, S., 2025. Predicting student academic path using machine learning: Systematic review. *2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, IEEE, 1–8. https://doi.org/10.1109/IRASET64571.2025.11008352

[7] Cortez, P., Silva, A., 2008. Using data mining to predict secondary school student performance. *Proceedings of the 5th Future Business Technology Conference (FUBUTEC)*, Porto, Portugal, 5–12

[8] Gull, H., Saqib, M., Iqbal, S. Z., Saeed, S., 2020. Improving learning experience of students by early prediction of student performance using machine learning. *2020 IEEE International Conference for Innovation in Technology (INOCON)*, IEEE, 1–4. https://doi.org/10.1109/INOCON50539.2020.9298266

[9] Hashim, A. S., Awadh, W. A., Hamoud, A. K., 2020. Student performance prediction model based on supervised machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*, 928(3), 032019. https://doi.org/10.1088/1757-899X/928/3/032019

[10] Soyoye, T. O., Chen, T., Hill, R., McCabe, K., 2023. Predicting academic performance of university students using machine learning: A case study in the UK. *2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE, 431–434. https://doi.org/10.1109/WI-IAT59888.2023.00071

[11] Alsalem, G. M., Sarhan, N., Hammad, M., Zawaideh, B., 2024. Predicting students' performance using machine learning classifiers. *2024 25th International Arab Conference on Information Technology (ACIT)*, IEEE, 1–5. https://doi.org/10.1109/ACIT62805.2024.10877102

[12] Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., Nawaz, R., 2020. Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104, 106189. https://doi.org/10.1016/j.chb.2019.106189

[13] Nabil, A., Seyam, M., Abou-Elfetouh, A., 2021. Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access*, 9, 140731–140746. https://doi.org/10.1109/ACCESS.2021.3119596

[14] Verma, B. K., Srivastava, N., Bharti, A. K., 2024. Advanced models for predicting student performance in higher education using machine learning and data mining techniques. *2024 13th International Conference on System Modeling & Advancement in Research Trends (SMART)*, IEEE, 600–605. https://doi.org/10.1109/SMART63812.2024.10882496

[15] Wang, J., & Yu, Y. (2025). Machine learning approach to student performance prediction of online learning. *PloS one*, *20*(1), e0299018.

[16] National Curriculum and Textbook Board (NCTB), 2012. National curriculum framework. Ministry of Education, Government of Bangladesh, Dhaka.

[17] Bangladesh Bureau of Educational Information and Statistics (BANBEIS), 2022. Bangladesh education statistics 2022. Ministry of Education, Government of Bangladesh, Dhaka.

[18] Biswas, S., Hassan, J., Tarek, L. R., 2024. Student Performance-BD [Data set]. *Kaggle*. https://www.kaggle.com/datasets/satayjit/student-performance-bd (Accessed 6 May 2025)

[19] Nahar, L., Chowdhury, M., 2019. Stream selection in secondary education of Bangladesh: Determinants and consequences. *International Journal of Education and Research*, 7(6), 123–134

[20] United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. New York: United Nations.