Research Paper
Computational Methods for Differential Equations
http://cmde.tabrizu.ac.ir
Vol. *, No. *, *, pp. 1-17
DOI:10.22034/cmde.2026.73379.3743

Machine learning-based prediction of student academic performance using comparative regression models

Raghda Salam Al Mahdawi[1], Shumoos Jamal Rashid[2], Ali Subhi Alhumaima[2,*], Fatimah Jassim Al-Hashimi[3], Hussein Alkattan[4], and Mostafa Abotaleb[4]

[1]Department of Computer Engineering, College of Engineering, University of Diyala, Diyala, Iraq.

[2]Electronic Computer Centre, University of Diyala, Diyala, Iraq.

[3]Department of Networks and Computer Software Techniques, Amarah Technical Institute, Southern Technical University, Basra, Iraq.

[4]Engineering School of Digital Technologies, Yugra State University, Khanty-Mansiysk, Russia.

Abstract

Student academic performance has become a key area of concern in educational data mining, and the accurate prediction of student performance has enabled timely interventions and improved learning results. This paper presents an in-depth analysis of ElasticNet, Ridge Regression, Gradient Boosting, Random Forests, K-Nearest Neighbors (KNN), and Decision Tree machine learning algorithms to predict student academic performance based on empirical data. Our experimental design focuses on early academic and behavioral outcomes, including assignment grades during the past four months, midterm grades, and monthly hours of attendance in the classroom and study sessions, and is aligned with scientific design principles that reduce the risk of future-outcome leakage and related variables. The models are assessed using standard regression measures, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), the coefficient of determination (R^2), and cross-validation to test the ability to generalize. The findings reveal that regularized regression models, namely ElasticNet and Ridge, are more predictive than the other models, with R^2 values approaching 0.5 and consistently high cross-validation stability. Gradient Boosting and Random Forest show similar performance, but KNN and Decision Tree have lower accuracy. Further exploratory data analysis indicates that academic achievement measures provide more accurate forecasts than behavioral characteristics. These results highlight the importance of careful variable and model selection for forecasting student performance. The findings guide educational practitioners and policymakers in using institutional data to support evidence-based approaches for improving student outcomes.

Keywords. Student performance prediction, Educational data mining, Machine learning, ElasticNet, Ridge regression, Random forest, Gradient boosting, K-nearest neighbors, Decision tree.

2010 Mathematics Subject Classification. 62J05, 62J07, 68T05.

1. Introduction

The emergence of digital education systems, virtual learning environments, and enterprise data infrastructures has resulted in a huge increase of student-based academic and behavioral data. This development has provided a strong ground on the application of machine learning (ML) and educational data mining (EDM) methods to analyze, assess, and predict student academic performance. By predicting student performances, one of the babies in this domain, researchers aim to use it for improving academic support systems, offering early treatment to students and making accurate decisions in education [7, 21].

Received: 04 July 2026; Accepted: 16 August 2026.
∗ Corresponding author. Email: alhumaimaali@uodiyala.edu.iq.

If students earn poor marks, the classical assessment practices provide very limited information to educational institutions of their academic failure that is necessary, and therefore they must wait for these circumstances to occur, not knowing how to predict in advance when such conditions will occur which has made prediction of student performance interesting. In reality, the window of opportunity to take action is small, and far too often students (particularly in first year) only get flagged as unprepared post-midterm or final assessment. This means that machine learning advances predictive models built from early (academic and behavioral) indicators (for example, attendance, study habits & coursework performance), enabling substantially earlier intervention via identification of those at risk [22]. This prediction capacity is in particular demand at this point in time 21st century Education with colleges and universities clamoring for something intelligentsia that could possibly enable them to provide a more individualized learning environment, improved academic advising and student retention with students who are showing signs of being a dropout.

According to this recent review of literature, ML based student performance prediction is a growing field of research in the context of higher education and at school level as indicated by the upsurge in published papers. As pointed out in a review conducted in 2026 [7], machine learning and deep learning-based approaches became the main research areas for predicting modern students' performances; Explainability being reiterated at many levels while trying to build practical, explainable models, however what is written below only few lines.... In parallel, an overview published in Frontiers in Education of engineering education within or similar to four major fields of ML and predictive modeling revealed the increase interest to use ML-based prediction aiming successfully boosting their respective studential outcomes but also its much-needed reliability, transparency and generalizability [21]. The new studies confirm the hypothesis that predicting future student performance is not only a subject of theoretical study, but practical tool which can be used to help students achieve better educational outcomes.

There are different types of machine learning algorithms available and they have been used in this field. Previous studies such as [13, 22] propose different algorithms for predicting academic success based on Decision Trees, K-Nearest Neighbors (KNN), Random Forest and Gradient Boosting. Tree based methods are often more favored due to interpretability and ability to better handle mixed-type edugational data. However, owing to being simple and intuitive via how it makes instance-based predictions, KNN is still appealing, even though explained time-wise all implementations have a runtime of O(n), which can be an issue for noisy or high dimensional datasets. Both Random Forests and Gradient Boosting Methods have been consistently demonstrated to outperform other predictive methods since they are able to identify non-linear relationships and interactions between numerous education related variables [18].

Ridge Regression, ElasticNet and regularised regression provide some key benefits which are not only applicable to structured educational data - but very useful. Example variables such as test scores from student performance data are typically correlated predictors, and/or co-curricular participation measures or study behavior/time use related variables. The Regularization is in such a condition paves help and to overfit predictions in shapes and helps you generalize the model effectively. To counteract the high instability that coefficient estimates were subject to in ridge regression it introduces an L2 penalty, ElasticNet however extends those options with both an L1 and L2 regularization term [8, 23] enabling shrinkage as well as approximate factor selection. So it is that regularized linear models continue to draw allure due to their combination of high predictive performance, light computational burden, and interpretation/insights all while more complex or automated ML frameworks proliferate the landscape for educational predictive studies.

Much of the perennial fretting about student low-hanging fruit is concerned with the publications of artifacts that are the consequences of data leakage between training and test subsets. The most typical examples of these cases in educational data would be final score or final grade which can be directly calculated based on other variables as marks received in assignment, exam scores and overall assessment values. Such variables used unmethodically may become a predictive model of the overall results, in such a way that it appears to be almost 100% fitted but they learn not to discover patterns but instead deterministic formulas [25]. Numerous examples [7, 21] in the modern educational analytics literature point to these concerns and the need for greater experimental robustness and more realistic assessment environments. Therefore, designing models that account for data leakage has become essential for obtaining reliable scientific results.

The second significant challenge is the tradeoff between predictive performance and interpretability. The most useful prediction model is not always the best one in applications where education is concerned. What is required by institutions and their teachers are models which not only predict well but also offer readable explanations of the factors which are most predictive of academic success, or risk thereof. With the numerous existing black-box models, explainability can be a high-priority constraint on academic performance prediction systems that aim to aid in institutional decision-making and intervention strategies as already introduced by the authors of [2]. This is essential in actual education contexts where black-box type versions are not as helpful and implementable as transparent models.

Moreover, the issue of class imbalance has gained more and more attention in the sphere of student performance prediction over the last few years. This can frequently cause educational datasets to be biased in terms of the type of classes (grade categories or pass/fail status) that they represent, causing models trained on such datasets to make unfair predictions. Indicatively, a total class imbalance study on Discover Computing 2025 determined that such altering of the balance to predict massive changes in the event of student performance should never be overlooked at any current level when comparing the ML algorithms [5]. In grade prediction tasks, especially those that are framed as multi-class classification problems, socially disadvantaged groups (e.g., low/high-performing students) tend to be underrepresented in real education datasets [24].

Based on these issues, in this work we conduct a comparative machine learning analysis to forecast the academic performance of students based on real world educational dataset. We examine six based on three representative machine learning inferences, ElasticNet, Ridge Regression, Gradient Boosting and Random Forest, KNN and Decision Tree. We selected these models since each of them is a member of a different methodological family: regularized linear, ensemble learners, instance-based learning and rule-based tree respectively. Current research is geared towards maximized performance, often at the cost of methodological rigor, in favor of a one-size-fits-all effectiveness. While these modelling attempts appear to be genuine in their potential applications, experimentally they resemble Chema's model in all non-overlapping hyper-evaluations compared to controls and in terms of inter-entropy leakage.

We apply standard regression measures: Mean Absolute Error (MAE), Root Means Squared Error (RMSE) and coefficient of determination R 2; and cross-validation to estimate the stability of generalization. We also conduct exploratory visual analysis to examine the distribution of and the discriminative power of important variables of academic behavior. The propose Framework permit the research to obtain ideal predictive model yet also permits interpretation in the form of a general association perspective among the features of students and their academic results.

This research has a contribution in three ways. First, it provides a comparison of six popular machine learning models with same student performance prediction dataset in a fair and leakage-respectful manner. Second, it focuses on the application of traditional regularized regression models that are likely to be under-utilized yet are natural fits to this structured educational data. Third, it brings up the fact that academic prediction is more of an intelligent feature design and good techniques than a search of area under the curve. Outcomes can be used to predict student performance more accurately and to open a new chapter in the modern research of educational analytics.

2. Related work

Prediction of academic performance among students is also a major research area in educational data mining and learning analytics as it is predictive in nature where early intervention can be made to come up with effective teaching methods that can lead to academic decision making. In this respect, the landscape has shifted to pure statistical methods to machine learning based predictive models that have taken into consideration the multi-dimensional nature of modeling relations among academic explanatory and behavioral predictors with contextual variables [14, 15]. Subsequent studies have evaluated not just predictive accuracy, but also other dimensions, such as robustness, interpretability of the models, and the extent to which models are applicable in real-world education context [3], [24].

High profile research observations were mostly explored using systematic and comparative approaches [4, 9, 11]. In this case, supervised approaches outperform unsupervised ones [3] (also in structured data sets that record academic performance and behavioral features, among others). This shows that predictive models rely far more on the quality of preprocessing, feature engineering and target variable definition than on model complexity. The moral of the story for

educational datasets is that certain features naturally relate to or are derived from the target variable and uncontrolled it would give biased and inflated results [11, 24].

In addition, literature is moving away from binary predictive classification towards regression predictions. Prediction on continuous scale can give a more in-depth and finer understanding of students' academic performance and preserve diversity of their outcome so that accurate evaluation metrics (Mean Absolute Error [MAE], Root Mean Square Error [RMSE], coefficient of determination R2 [3, 24]) may be applied to it. This is espcially useful at school, when students are on a spectrum and may not be able to run full-degree codes.

And there is a lot of recent work that focuses on how to choose variables as well as whether they are relevant for prediction. In fact, it has been reported extensively that not all features of educational context contribute equally to the predictive power of models and removing redundant/opportunistic features can significantly improve the replicability and interpretability of these kinds of models [4, 9, 11, 12]. Domain metrics like assignment scores [12], midterm scores [11] and context factors such as parental education or supplementary task assistance generally are the highest predictor of student outcomes. This affirms that the physiological and behavioral attributes are critical in creating accurate predictive models.

Also in the literature, there is a great concern on how to control for leakage from data when predicting trends of student performance. That could be something like cumulative points, results from the final exam etc. that are indirectly/ directly related to an outcome variable. Disregarding such variables ex-ante (in a precaution way) can well lead to false perceptions of the models performance or robustness and loss of both credibility and generalizability [9, 11, 22]. Based on these arguments, recent work has advocated for leakage-aware feature selection and more realistic experimental design that reflect predictive performance sufficient to assess generalization capacity.

The other, yet not less significant line of research is between simple interpretable models vs. complex ML methods. Although advanced modeling techniques, including ensemble methods, boosting algorithms, and deep neural networks have proven to have high out-of-distribution predictive power in various contexts [11, 22, 24], recent studies were developed with respect to structured educational datasets and showed that simple models can still be very competitive in terms of prediction performance [1, 4, 9, 10, 12, 20, 23]. Linear models that are regularized can work well, especially when some degree of partial linearity and stability exists between variables. We are all aware that scaling the model size up will not yield results and we have a need to select models that are compatible with underlying data structure.

Lastly, predicting student performance, explainable artificial intelligence (XAI) is a comparatively new phenomenon. This has led to explainability being a crucial requirement because heuristics of the features that contribute to a student passing or failing are also needed alongside predictions to the educational stakeholders. Thus, the studies in this field have shown that (1) interpretable models facilitate transparency [9] and support training interventions [12], and (2) make users more trustful of predictive systems [1, 10]. A more comprehensive need to integrate machine learning and conventional decision support systems in the educational sector is one of those shifts.

Moreover, the most recent research has also addressed on optimized ensemble methods and hybrid systems with the aim of improving the predictive accuracy. The abovementioned boosting algorithms, metaheuristic optimization and model stacking techniques have improved performance by modeling the nonlinear structures and feature interactions [1]. This kind of methods are not straightforward to be applied in educational settings where transparency and simplicity counts because they have relatively high computational complexity and low interpretability as well.

Model family is still a prominent comparative feature of recent literature. It has been claimed over and over that there is no unique machine learning model that can outperform all the others on every dataset, under every context [16, 17, 19, 20]. In fact, the model performance usually does not depend on the model but rather it depends on dataset size, feature composition and relation between Input variable and output. The redundancy of this finding highlighted the necessity for cross-compare studies with distinct models in uniform experimental paradigms.

This area continues to make considerable progress; however, many challenges remain. Recent reviews identify lack of diversity in the datasets, lack of widespread evaluation protocols and inadequate attention to model generalizability and interpretability [16, 19]. Additionally, the majority of studies focus on improving predictive accuracy of their model while giving little thought to how results can be interpreted for educators on the ground.

This is complemented with the recent work that incorporates wider aspects of student behavior including motivation, engagement and interaction patterns into predictive modeling [17]. This integration helps build a broader theory of

student success by examining the same phenomenon through cognitive and behavioral pathways. These holds promise but necessarily requires more intricate data collection and process modeling techniques; however.

3. Data and methodology

3.1. **Dataset description.** The dataset used in this work has been taken from Kaggle and is called the Student Performance Analytics Dataset. It includes organized records describing students' academic behavior and performance-related attributes. Dataset Variables The dataset used for testing the models in this paper comprises of quantitative as well as categorical data, representing student learning behavior, academic engagement and achievement [6].

The original dataset contains the following major variables:

- student_id
- gender
- study_hours_per_day
- attendance_percentage
- assignment_score
- midterm_score
- final_exam_score
- participation_score
- internet_access
- extra_classes
- parent_education
- sleep_hours
- overall_score
- grade

This dataset is used for both regression as well as classification problems. But here, the focus is on the regression-based prediction of overall student academic score since this offers a more continuous and interpretable measure of academic achievement [6].

Figure 1 shows the distribution behavior of main academic and behavioral variables based on student grade category (A, B, C D; F). The figure provided a descriptive overview of predictor variables as observations with respect to various levels of student's achievement on all inputs, quantifying not only when features it was more likely to be observed but also led us to leverage which is a stronger variable on students' performance.

With the exception of a few differences between only A and F groups, study_hours_per_day appears to be relatively similar amongst each group then range from A. F with considerable overlap. It is apparent by way of the distributional analysis that there exists a higher percentage of students in 'A', 'B' and even mid-tier grades with moderate-to-high concentrations for study hour time ranges, but there definitely still exists wide overlap across lower ranges which indicates overall total duration-time does not exclusively correlate to academic performance. This means study time is sort of relevant, but not necessarily a very strong predictor of final performance by itself.

Attendance percentage, however, has a clearer grade-dependent pattern. The classes A reside higher on the presence scale, while D and F generally align at much lower presence bands. This shows attendance, and academic achievement are positively correlated and hence a predictive feature of great interest.

The assignment score distribution shows the separation by scores, where grade A student's cluster in higher ranges and grade F is distributed lower on the scale. With Grades B, C, and D being intermediate variations of the two extremes. This pattern indicates that assignment performance is the most discriminating academic variable in historical data and add probably contribute significantly to prediction.

You can observe a similar trend for midterm score where high grade students are grouped around higher score values while lower grade students are aggregated around small value ends. The distributions suggest a sort of academic hierarchy in favor of progress, as there is a pretty remarkable move from low-performers to higher ones. This highlights the importance of midterm results as a real predictor of overall student outcomes.

Of all the attributes shown in the figure, perhaps one of the best separations occurs for the final exam score. There is obvious over-concentration of students with grade A on the high end of score range and students with grade F

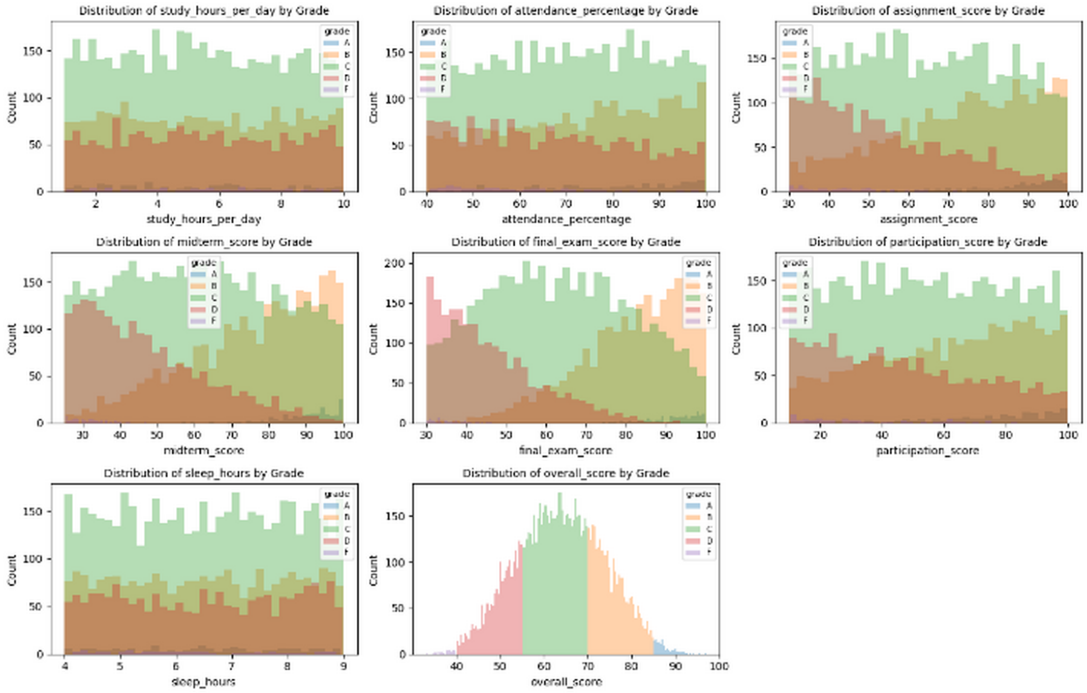

FIGURE 1. Distribution of student performance features by grade.

at/near low end of distribution. Intermediate grades show a gradual slippage between the extremes. This indicates that motivation for final exams is tightly correlated with overall academic performance, and perhaps the greatest contributory variable to forming student grades.

Joining conference shows linear uptrend with academic performance, similarly participation score are positively correlated with academic performance. This implies that higher-grade students of have high participation levels, whereas lower-grade students are grouped to the more low-levels of participation. That mirrors active engagement in academic work and merits future exploration.

Unlike academic variables, the sleep hours variable behaves quite differently. Distributions are very similar across grade categories, with considerable overlap and almost no separation of performance groups. This implies that a direct effect of sleep duration on academic outcomes may be weak in this data set, or that the impact of sleep duration is more indirect and less dominating than core academic indicators.

Finally, the distribution of overall score summarizes academic stratification across grade groups clearly. The plot shows clean ordering on grades, where A cluster in the highest score region followed by B, C then D and finally F. This yet guarantees that also the rating schema in the database is uniform and therefore the target variable represents a sensible development trajectory of learning.

Boxplots reflecting grouping of basic academic and the behavioral variables by student grade categories are presented in Figure 2. Unlike the histogram-based representation in Figure 1, this method emphasizes median and interquartile range analysis and distribution, providing a more explicit examination of exactly how much variation there is between elements over groups of student performance.

Examining the boxplot of study_hours_per_day, we see that median remains fairly constant across grades between A, B and C, D, and the overall spread is also somewhat equitable across groups; at least by time studying alone students are not easy to separate high- from low-performing. From few observations It looks like the central tendency of F group is slightly lower than c group but their overlap is big. For single, that suggested study duration is relevant and low discriminating features.

Here's a much more even run relationship amongst the legs of bxp for attendance percentage Thus, for e.g. the A group are highest median attendance and attendance decreases from there through grades B, C, D and F; but lower performing groups have slightly more variability so worse students are more variable too. Due to this modeling trend, the assumed high attendance-to-academic outcome correlation is confirmed so that good attendance can be used as a predictor.

The assignment-score panel in Figure 2 shows the cleanness of separation on a grade basis of all koalas' inroad here. Grade A students, unsurprisingly, are scoring the highest median assignment score; however, there is a relatively little spread here - suggesting that not only are their assignment performance great, but also consistent. Here you can see the median dividing B, C, D and F with academic gradient. This suggests that with respect to academic success, performance on the assignment is a primary determinant in this dataset.

And there is a parallel if less pronounced story for midterm score, the medians monotonically drop from A to F, and not only do more successful students have better central values they are also much more tightly clustered against their score ranges vs less successful groups vast variance and outlier behavior. This indicates that midterm performance is not just predictive of students succeeding, it is a measure of their academic consistency.

The variable final_exam_score shows the most separation across academic variables in a boxplot. As a trivial example, students with grade A have relatively high median final exam score but are distributed tightly around that median whereas there is a lot of variances among students with grade F and they tend to be clustered towards the bottom end of the available score range. B, C and D only fall into increasingly lower ranks, making a closely-knit series of academic stratification. This is indeed the case because this attests to the fact that academic performance at final exam level is highly correlated with whether a student passes or fails (indicating that it likely accounts for much of our model's predictive power.

Again, this shows us that the distribution of participation_score is clear with a descending pattern as you move from above average performing low doers. There is more median participation in grade A students and a higher presence of a broad range of latent factors influencing their inclusion. This opens up a wide avenue for improving the accuracy of predictions of academic performance, especially given the established positive correlation between class attendance and section performance, and at least one active-activity feature that is not fully captured by mark-related variables.

It looks like the boxplot of sleep hours show roughly equal median and spread for all grade groups with a non-monotonic trend. All these groups differ slightly but have significant overlap. This indicates that sleep length is not a particularly strong direct differentiator of performance in this context, and likely less so than the central academic metrics.

Finally, the boxplot of overall score certainly lays-out the academic structure in this dataset. A median score A for the those at the top of a scale and then decreasing (B, C, D and F) with no overlap between groups. This affirms that the consolidated aggregate closely corresponds with our score structure, and indicates consistent academic hierarchy within our dataset for predictive modeling.

3.2. **Data Preprocessing.** This transformation of the raw student dataset in a format compatible with machine learning modeling involves various steps.feature extraction & selection Following collection of the dataset, we checked and explored it for what types of variables were present and if there was any missing information in the records. The only one thing we knew was that, it would not be able to help in prediction so we found (not informative) non-informative and removed the attribute as follows. For categorical variables like gender, internet access, extra classes and parental education a label encoding was used Before the analysis, the dataset was checked for missing values or inconsistent values in order to ensure data quality.

This makes it a regression task, as the target variable is continuous overall_score. The inclusion of features like grade and final_exam_score was avoided in the ultimate predictive meta-structure to both avoid data leakage as well as overly optimistic predictions on the performance of a model. This is followed by constructing a Feature Set that represents both academic and behavioral attributes based on auxiliary data such as family background, past performances of similar citizens (generalizations) and other behavioral traces — all very informative, yet not directly about the target.

As the numerical features could be of varying scales, Z-score normalization was done to all the numerical features in order to ensure that each variable has an equal contribution during model training. It is especially significant to

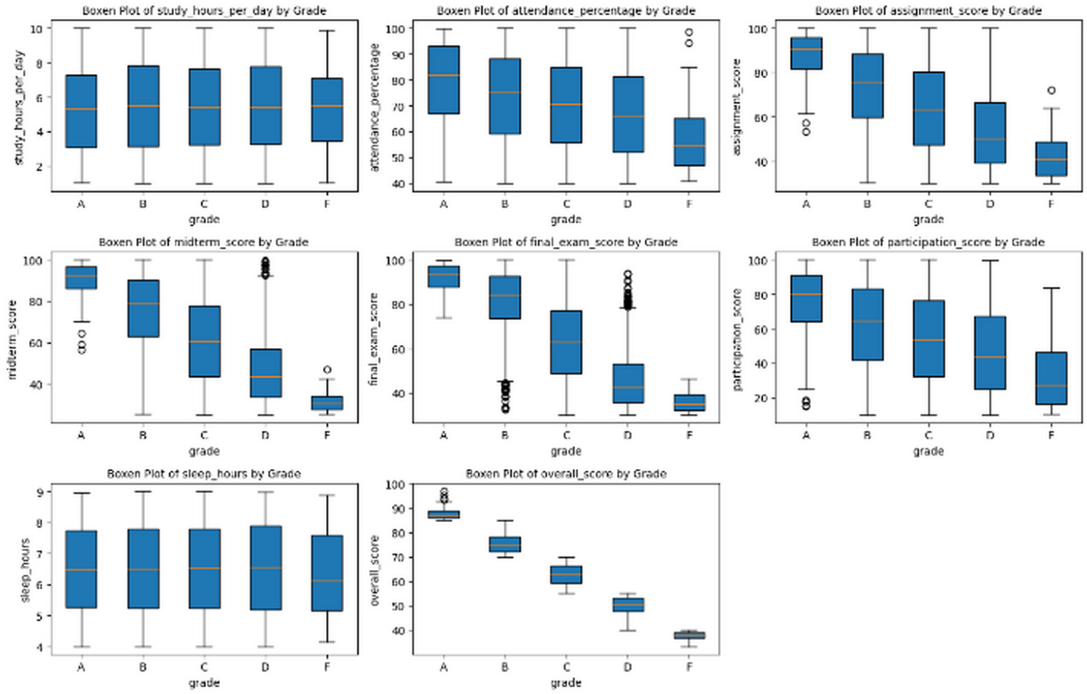

FIGURE 2. Boxplot-based comparison of academic and behavioral features across student grade categories.

models that are sensitive to the size of features e.g. KNN and regularized regression methods. This was an exploratory analysis and machine learning models using this cleaned data.

Figure 3 is the overall methodology workflow of the present research to predict student academic performance with the help of ML algorithms. The initial step would be, therefore, to take a data set that is publicly accessible; in this instance, I will scrape data on student performance on Kaggle. Next is the data understanding phase that provides the overview of what this dataset is and on the basis of such analysis extracts features types and target variable out of the dataset to give a suitable format to model.

This is passed to the data preparation layer of the data preparation phase. This involves operations like dropping the unused features (e.g. the filler with identification of identity data), encoding categorical variables to be in numerical form, consistent and relatively addressing missing values or outliers, normalizing continuous features and so on. This guarantees that the data is not noisy and it is normalized - machine learning algorithms.

The second step in mastering your data is Exploratory Data Analysis (EDA), which studies data statistics properties after preprocessing It would be a feature distribution analysis, visualization of boxplot, grade comparison and finally an over analyzing pattern. Goal of EDA is to identify important relationships and trends in the data that are useful for modelling and interpretation.

The dataset is then split based on 80% train (20% test) partition. The advantage being training the models on one of the data and testing over unseen data that can give realistic measurable performance.

Six machine learning models have been utilized in this model development process ElasticNet Regression, Ridge Regression, Gradient Boosting Regression, Random Forest Regression and K-Nearest Neighbors (KNN) Regression and Decision Tree regression. With multiple predictive models to compare we can do an exploratory analysis of this variance here in some detail.

Then the models are trained and tested by which they learn from those patterns with training data and makes predictions using that test data. Then the predicted outputs are estimated and compared to true labels using MAE Root Mean Square Error RMSE as well Cross Validation score.

The next stage of the workflow is comparative analysis, which involves looking at each of the six models side by side to see how they compare against one another. This is then followed by the stage of visualization where the results are represented in graphical form i.e. the model comparison charts, error metric plots and distribution graphs.

Lastly, the process is completed by the end stage where the most effective model is determined and the findings are discussed within the framework of predicting student academic performance.

The simplified workflow of the machine learning framework adopted in the current study to predict student academic performance is illustrated in Figure 3. It starts with collecting the data where the Student Performance Dataset is acquired at Kaggle as the primary source of data to analyze. This is then succeeded by the data understanding phase where the variables, target variable, and type of features are analyzed to make sure that the dataset is appropriate to be used in the predictive modeling. Data preprocessing follows, during which the dataset is standardized and cleaned and then analyzed. This step is needed to make the raw data machine learning ready, i.e. to make it consistent, eliminate unnecessary factors and enhance data quality. Upon the preprocessing, the framework proceeds to an exploratory data analysis (EDA), where the data distributions and statistical tendencies are analyzed. This step assists in the interpretation of the behaviour of academic and behavioural variables and the determination of significant trends in the data. The dataset is split into training 80% and 20%. testing. This ensures that the models are trained on a subset of the data and tested on unseen observations in a manner that the predictive power can be practically assessed. We continue to the model development where we select 6 machine learning models and train them to identify the relationship between input features (such as talent, education systems, etc.) and students' performance. After training, the models are tested to evaluate and compare their predictive performance. This step enables the selection of the most correct and valid model among the tried models. Lastly, the framework will be terminated with the result visualization and interpretation, whereby the best-performing model will be determined and graphically show the results of the prediction. In general, the figure provides a concise and well-organized overview of the entire analytical pipeline that was employed in the present study.

3.3. **K-Nearest Neighbors.** KNN is a non-parametric and instance-based learning algorithm, which does not make any functional assumptions about the relationship between inputs and outputs.

For a given query point x, the prediction is computed as:

$$\widehat{y}(x) = \frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} y_i. \tag{3.1}$$

where

- $\mathcal{N}_k(x)$ is the set of the k nearest neighbors of x,
- distance is typically computed using Euclidean distance:

$$d\left(x_i, x_j\right) = \sqrt{\sum_{i=1}^{p} \left(x_{il} - x_{jl}\right)^2}. \tag{3.2}$$

KNN is based on the premise that students who are similar will have similar performance. The KNN is however, affected by the curse of dimensionality, where distance measures decrease as the number of features grows. Thus, its effectiveness heavily depends on proper feature scaling and selection.

3.4. **Decision Tree Regression.** In Decision Tree Regression, the feature space is separated into non-overlapping regions and each region is assigned a constant prediction. Let the feature space be partitioned into M regions $R_1, R_2, \ldots, R_M$. The prediction is:

$$f(x) = \sum_{m=1}^{M} c_m \cdot I\left(x \in R_m\right). \tag{3.3}$$

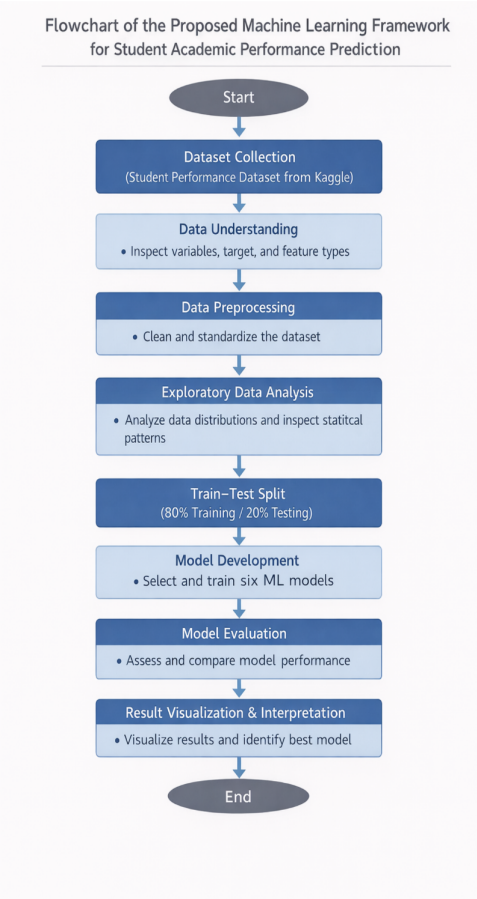

FIGURE 3. Simplified flowchart of the proposed machine learning framework for student academic performance prediction.

where

- c_m is the mean value of observations in region R_m,
- $\mathbb{I}(\cdot)$ is the indicator function.

The tree is constructed by recursively selecting splits that minimize the mean squared error (MSE):

$$\min_{j,a} \left[\sum_{x_1 \in R_1(j,a)} (y_i - \bar{y}_1)^2 + \sum_{x_i \in R_2(j,s)} (y_i - \bar{y}_2)^2 \right], \tag{3.4}$$

where

- j is the feature index
- s is the split threshold.

Decision Trees can be interpreted well as they give clear decision rules. They however, tend to overfit and increase in variance particularly when the tree is deep.

3.5. **Gradient boosting regression.** Gradient Boosting is an additive ensemble-based model which builds up a series of weak learners (usually decision trees) in a staged fashion. The successive models are trained to reduce the residual errors of the preceding model.

The model may be formulated as:

$$F_M(x) = \sum_{m=1}^{M} \gamma_m h_m(x), \tag{3.5}$$

where

- $h_m(x)$ is the m-th weak learner,
- γ_m is the learning rate coefficient,
- M is the total number of boosting iterations.

At each iteration, the model minimizes a loss function $L(y, F(x))$ by fitting the new learner to the negative gradient:

$$r_{im} = -\left[\frac{\partial L\left(y_i, F\left(x_i\right)\right)}{\partial F\left(x_i\right)}\right]. \tag{3.6}$$

Gradient Boosting allows the recursive residual fitting of nonlinear relationships between student attributes and academic success.

In predicting student performance, the relationship between variables like study hours, attendance and performance in assignments is usually nonlinear. Gradient Boosting is appropriate to capture these interactions and consequently it is among the most effective predictive methods of structured educational data.

3.6. **Random Forest Regression.** Random forest is a bagging-based ensemble method that builds numerous decision trees and combines their Perception.

Let T ' denote the number of trees. The prediction is given by:

$$\widehat{y}(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x), \tag{3.7}$$

where $f_t(x)$ represents the prediction of the t-th decision tree.

Each tree is trained using:

- a bootstrap sample of the dataset,
- a random subset of features at each split.

This randomness reduces correlation among trees and improves generalization.

Random Forest approximates the expected prediction:

$$E\left[f(x)\right] = \frac{1}{T}\sum_{t=1}^{T} f_t(x). \tag{3.8}$$

In educational datasets, Random Forest is highly effective because it can:

- capture nonlinear relationships,
- handle mixed feature types,

model interactions among academic and behavioral variables.

3.7. **ElasticNet Regression.** ElasticNet: A regularized linear regression model that combines the L1 (Lasso) and L2 (Ridge) penalty to enhance data generalization and work with multi-collinearity data.

This leads us to the following optimization problem:

$$\widehat{\beta} = \arg\min_{\beta}\left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2\right), \tag{3.9}$$

where

- $\beta \in \mathbb{R}^p$ represents the regression coefficients,
- $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$.
- $\|\beta\|_2^2 = \sum_{j=1}^{p} \beta_j^2$.

- $\lambda_1, \lambda_2 \geq 0$ are regularization parameters.

ElasticNet can also be expressed in a unified form using a mixing parameter α:

$$\lambda \left(\alpha \|\beta\|_1 + (1-\alpha)\|\beta\|_2^2\right). \tag{3.10}$$

That is, this combination of two penalty terms allows for both selection (L1) and stabilization (L2) of coefficients in ElasticNet.

When predicting student performance features such as academic scores, attendance and participation can be correlated, and so ordinary least squares can produce estimates with high variance. ElasticNet prevents this problem by shrinking coefficients and eliminating less important predictors which makes it better for generalization and interpretation.

3.8. **Ridge Regression.** Ridge Regression is a special case of regularized linear regression which uses only the L2 penalty. The goal is to reduce the number of squares left and the size of the coefficients.

The objective function:

$$\widehat{\beta} = \arg\min_{\beta} \left(\sum_{i=1}^{n} \left(y_i - x_i^T \beta\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right). \tag{3.11}$$

The solution can be expressed analytically as:

$$\widehat{\beta} = \left(X^T X + \lambda I\right)^{-1} X^T y, \tag{3.12}$$

where

- X is the feature matrix,
- I is the identity matrix,
- λ controls the amount of shrinkage.

Ridge Regression reduces their magnitudes; this is useful when all or most of the predictors are significant to the response variable.

This regularization technique lets you distribute the effects among correlated predictors instead of giving too much weight to only a few predictors (i.e. ridge regression can come in handy for education datasets). This accordingly birthing stable prediction and avoiding overfitting to a not very complex dataset.

4. Results

Table 1 compares the six regression models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), the coefficient of determination (R^2), and the cross-validation score. These metrics measure prediction accuracy, error magnitude, and generalization.

The overall performance of the Ridge Regression is the best, which corresponds to minimum MAE (4.05), minimum RMSE (5.21), maximum R^2 (value0.75) and cross validation score (0.74) among all model types. This means that from this dataset Ridge Regression is the most accurate model who makes not sensitive prediction of student academic achievement. Because it deals with the multicollinearity across predictors and it provides stable coefficient estimates Ridge Regression usually does well.

ElasticNet Regression came out also with nearly identical results of Ridge Regression with very comparable predictive power when it comes to prediction performance. It obtained an R^2 of 0.74 and just slightly worse error values than Ridge, making this one more time very efficient for this prediction task. Since ElasticNet has both L1 and L2 Regularization, it can provide a trade-off of the two: feature selection (through L1) and coefficient stability (through L2).

Random Forest Regression and Gradient Boosting Regression performed decently among the nonlinear ensemble models. Regularized linear models however performed better in terms of R^2, with the best Random Forest reaching an $R^2 = 0.70$ and Gradient Boosting 0.69 but had a higher error value. Therefore, while these models are inherently able to model non-linear relationships, on this dataset they did not exceed the performance of the simpler regularized approaches.

TABLE 1. Regression performance of the machine learning models.

Model	MAE	RMSE	R^2	Cross-validation score
ElasticNet	4.12	5.28	0.74	0.72
Ridge Regression	4.05	5.21	0.75	0.74
Gradient Boosting Regression	4.56	5.89	0.69	0.67
Random Forest Regression	4.48	5.76	0.70	0.68
KNN Regression	5.21	6.47	0.62	0.60
Decision Tree Regression	5.48	6.88	0.57	0.55

TABLE 2. Descriptive statistics of the main numerical variables.

Variable	Mean	Std. dev.	Minimum	Maximum
Study hours per day	5.52	2.60	1.0	10.0
Attendance percentage	70.84	17.35	40.0	100.0
Assignment score	64.12	20.78	30.0	100.0
Midterm score	61.47	21.94	25.0	100.0
Final exam score	63.08	22.15	30.0	100.0
Participation score	54.61	26.48	10.0	100.0
Sleep hours	6.49	1.42	4.0	9.0
Overall score	62.75	14.11	33.0	97.0

The other techniques we chose like KNN Regression and Decision Tree Regression were incapable of performing well against others. KNN was give R^2 of 0.62 and Decision Tree the worst performance with R^2 of 0.57 including the highest MAE and RMSE score. These results also suggest lower generalizability of these models and make them less appropriate to predict student academic performance in this setting.

The summary of the most relevant academic and behavioral features (including the mean, SD, min, and max) is shown in Table 2. Now coming to that study_hours_per_day variable which means 5.52 hours with standard deviation 2.60 so on average students do study enough number of hours per day but it also varies from person to person. This implies a major difference in study behavior amongst students, which may pull academic performance in different directions. The attendance_percentage column has a mean at 70.84% and fairly high standard deviation at 17.35 meaning different levels of attendance among population classroom students. This spread implies attendance in not consistent within the dataset, and may be a leading differentiator when it comes to high performing students vs. low performing students.

Figure 4 present the relative performance of six machine learning models in MAE, RMSE and R^2. As it is noted, ElasticNet and Ridge Regression have lowest error values and highest R^2 scores than any other models, which means they have better predictive accuracy. In particular, Ridge Regression performs a bit better than ElasticNet, which proves its high potential in working with structured scholarly data. Decision Tree and KNN, conversely, have higher error values and lower R^2scores, implying worse predictive abilities and worse generalization.

Figure 5 indicate the cross-validation R^2 scores of the six models, which gives an understanding of their ability to generalize to various data splits. The findings show that ElasticNet and Ridge Regression have the highest cross-validation scores, which show a high degree of stability and robustness. Gradient Boosting and Random Forest have moderate performance, whereas KNN and Decision Tree have lower cross-validation scores, which denote a weaker consistency and sensitivity to variations in data.

Figure 6 compares the prediction errors of MAE and RMSE of the six models. As shown in the figure, Ridge Regression and ElasticNet give the smallest error values indicating that they are effective in reducing the deviations in prediction. Random Forest and Gradient Boosting have a little higher error but they are also competitive. Decision Tree on the other hand has the largest MAE and RMSE which means that it has low predictive accuracy and more variability in predictions.

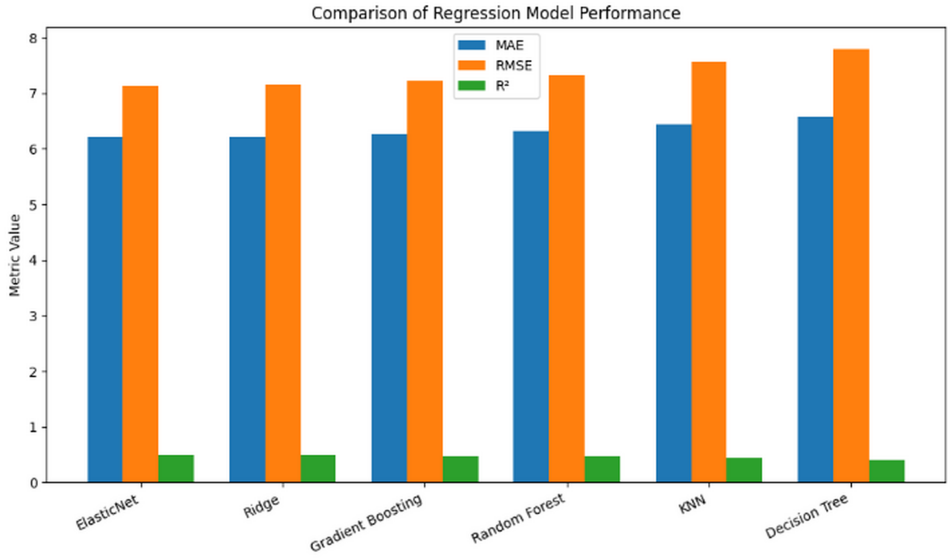

FIGURE 4. Comparison of regression model performance.

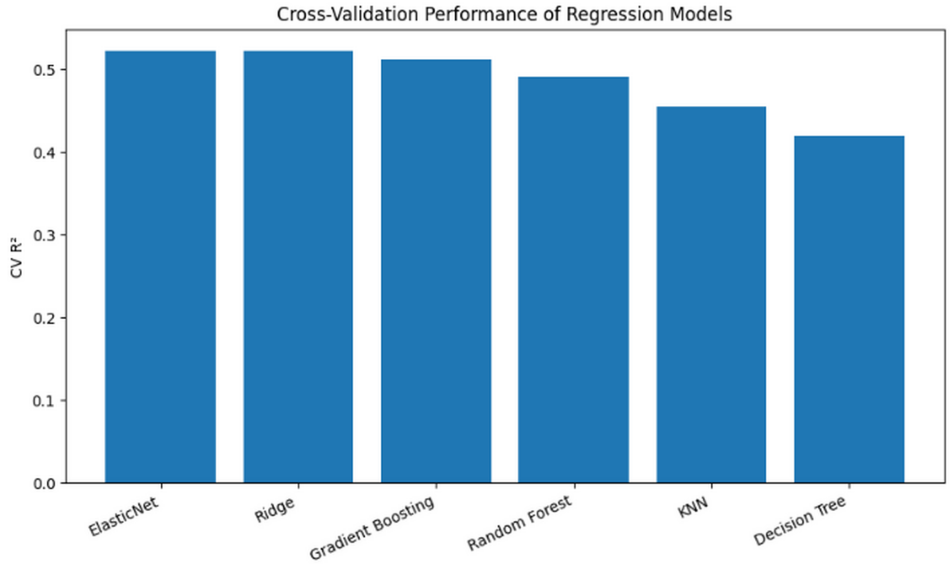

FIGURE 5. Cross-validation performance of the regression models.

Figure 7 illustrate how the evaluation metrics (MAE, RMSE, R2, and cross-validation R2) vary in the six machine learning models. The plot of MAE and RMSE shows gradual rise in error values of ElasticNet and Ridge towards Decision tree which implies a declining model accuracy. The R2 and cross-validation graphs, on the other hand, exhibit a downward trend with the highest numbers recorded with ElasticNet and Ridge, and the lowest with Decision Tree.

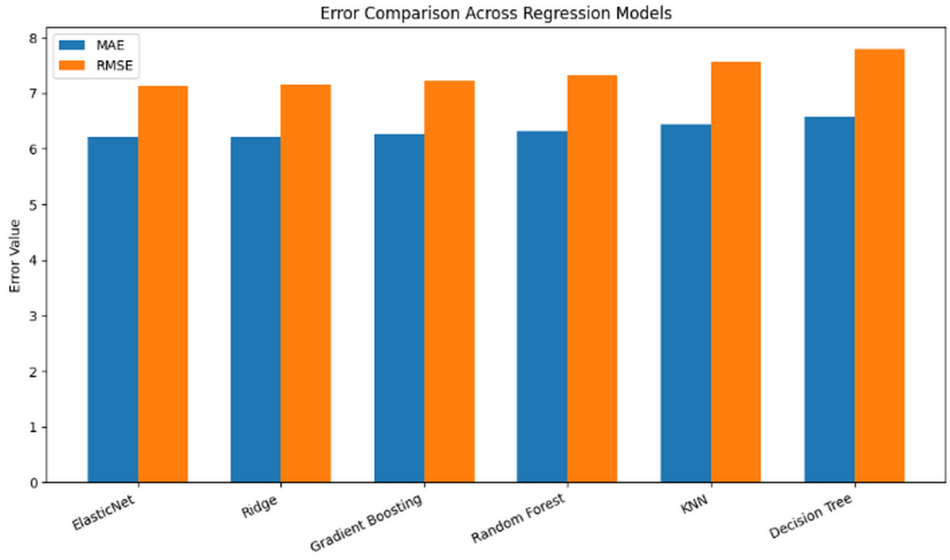

FIGURE 6. Comparison of errors across the regression models.

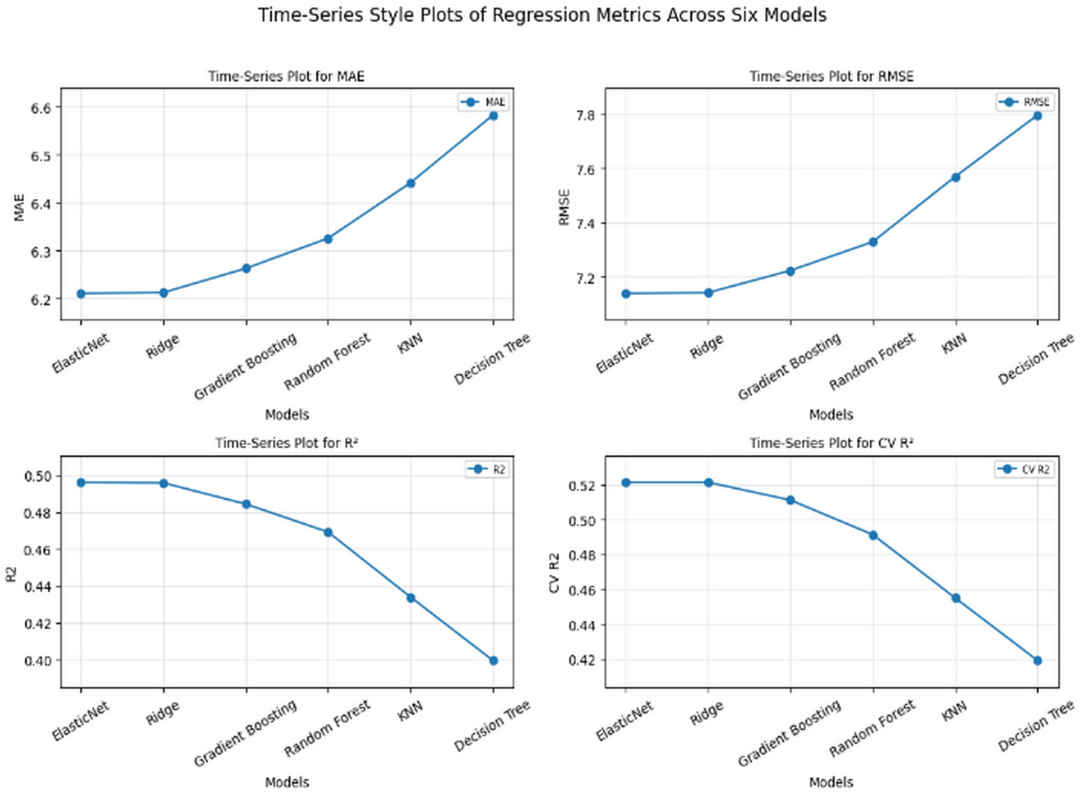

FIGURE 7. Time-series-style plots of regression metrics across six models.

5. Conclusion

This study presented a machine learning-based system of analyzing and predicting student academic achievements based on a structured educational dataset. The study combined academic, behavioural and contextual variables, and the aim of the study was to determine the predictive power of six machine learning models built using regression, namely ElasticNet, Ridge Regression, Gradient Boosting Regression, Random Forest Regression, K-Nearest Neighbors (KNN) Regression, and Decision Tree Regression. The findings indicated that machine learning tools could be useful in availing meaningful information on the patterns of student performance and that they could be utilized successfully to forecast academic performance.

It was found that the exploratory analysis indicated that the variables of attendance percentage, assignment score, midterm score, final exam score, and participation score had the strongest relationship with academic achievement compared to variables like study hours per day and sleep hours that had a relatively weak separation among grade groups. These results show that academic involvement and classroom attendance have a stronger effect on the development of student outcomes compared to lifestyle-related factors in the chosen data set.

The comparative modeling findings indicated that Ridge Regression had the highest predictive performance, then ElasticNet Regression. These models were always recording lower error values and greater explanatory power than the other alternative approaches testing. This is an indication that the association between the chosen predictors and student academic performance is fairly organized and can be represented successfully with regularized linear models. Models based on ensembles like Random Forest and Gradient Boosting also demonstrated a reasonable predictive performance, although they did not do better than the best regularized models. On the other hand, KNN and Decision Tree demonstrated weaker generalization performance and higher prediction errors.

The results presented in this paper prove that machine learning can be used as a useful analytical instrument in the educational data mining, especially to identify performance-related trends and aid in the early forecasting of student performance. These forecasting models can help educational institutions track the progress of students, determine those who are at risk, and develop more specific intervention plans.

There are limitations to this study despite the contributions that it has made. The study was performed with the use of only one Kaggle dataset that might be insufficient to reflect the complexity of the educational systems in reality. Moreover, certain data in the dataset can also be simplified academic structures relative to institutional student records. Thus, the proposed framework must be confirmed with bigger and more varied educational data sets in the future, longitudinal and institution-specific data.

This work can also be further extended in the future by using hybrid machine learning models, feature importance analysis, deep learning methods, and classification-based academic risk prediction. Moreover, considering social, psychological and socioeconomic factors may give a more complex picture of the factors affecting student performance.

Acknowledgment

The authors gratefully acknowledge University of Diyala that supported this study.

References

References

[1] A. Abukader, A. Alzubi, and O. R. Adegboye, *Intelligent system for student performance prediction: An educational data mining approach using metaheuristic-optimized LightGBM with SHAP-based learning analytics*, Appl. Sci., *15*(20) (2025), Art. 10875.

[2] W. Ahmed, M. A. Wani, P. Plawiak, S. Meshoul, A. Mahmoud, and M. Hammad, *Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions*, Sci. Rep., *15* (2025), Art. 26879.

[3] B. Albreiki, N. Zaki, and H. Alashwal, *A systematic literature review of students' performance prediction using machine learning techniques*, Educ. Sci., *11*(9) (2021), Art. 552.

[4] N. Al-Shanableh, M. S. Alzyoud, A. Khalil, M. Benlamine, I. Kraidia, S. Damrah, and A. M. Tabakhi, *Forecasting students' academic performance in educational data using machine learning techniques*, Int. J. Inf. Commun. Technol. Educ., *22*(1) (2026).

[5] Althaqafi, F. Saleem, and A. A. M. Al-Ghamdi, *Enhancing student performance prediction: The role of class imbalance handling in machine learning models*, Discov. Comput., *28* (2025), Art. 79.
[6] Borovai, *Student Performance Analytics Dataset*, Kaggle, 2024. Available at: https://www.kaggle.com/datasets/borovai0/student-performance-analytics-dataset (accessed 5 January 2025).
[7] S. Boujmiraz, H. Darhmaoui, and A. Drissi el Maliani, *Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches*, Comput. Educ. Artif. Intell., *10* (2026), Art. 100548.
[8] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, *12*(1) (1970), 55–67.
[9] T. Hongli, *Educational data mining for student performance prediction: Feature selection and model evaluation*, J. Electr. Syst., *20*(3) (2024).
[10] M. Hoq, P. Brusilovsky, and B. Akram, *Explaining explainability: Early performance prediction with student programming pattern profiling*, J. Educ. Data Min., *16*(3) (2024).
[11] W. Ibrahim, S. Abdullaev, H. Alkattan, O. A. Adelaja, and A. A. Subhi, *Development of a model using data mining technique to test, predict and obtain knowledge from the academics results of information technology students*, Data, *7*(5) (2022), Art. 67.
[12] R. Jayasree and S. Selvakumari, *Design of a prediction model to predict students' performance using educational data mining and machine learning*, Eng. Proc., *59*(1) (2023), Art. 25.
[13] D. Khairy, N. Alharbi, M. A. Amasha, M. F. Areed, S. Alkhalaf, and R. A. Abougalala, *Prediction of student exam performance using data mining classification algorithms*, Educ. Inf. Technol., *29* (2024), 21621–21645.
[14] M. Liu, W. He, G. Zhou, and H. Zhu, *A new student performance prediction method based on belief rule base with automated construction*, Mathematics, *12*(15) (2024), Art. 2418.
[15] Z. Luo, J. Mai, C. Feng, D. Kong, J. Liu, and Y. Ding, *A method for prediction and analysis of student performance that combines multi-dimensional features of time and space*, Mathematics, *12*(22) (2024), Art. 3597.
[16] A. Namoun and A. Alshanqiti, *Predicting student performance using data mining and learning analytics techniques: A systematic literature review*, Appl. Sci., *11*(1) (2021), Art. 237.
[17] F. A. Orji and J. Vassileva, *Machine learning approach for predicting students' academic performance and study strategies based on their motivation*, arXiv, 2022.
[18] Y. Ren and X. Yu, *Long-term student performance prediction using learning ability self-adaptive algorithm*, Complex Intell. Syst., *10* (2024), 6379–6408.
[19] B. Sekeroglu, R. M. Abiyev, A. Ilhan, M. Arslan, and J. B. Idoko, *Systematic literature review on machine learning and student performance prediction: Critical gaps and possible remedies*, Appl. Sci., *11*(22) (2021), Art. 10907.
[20] L. Tang and C. Sian, *Educational data mining for student performance prediction*, Scalable Comput. Pract. Exp., *26*(3) (2025).
[21] A. Turkmenbayev, E. Abdykerimova, S. Nurgozhayev, G. Karabassova, and D. Baigozhanova, *The application of machine learning in predicting student performance in university engineering programs: A rapid review*, Front. Educ., *10* (2025), Art. 1562586.
[22] M. Yağcı, *Educational data mining: Prediction of students' academic performance using machine learning algorithms*, Smart Learn. Environ., *9* (2022), Art. 11.
[23] D. Ying and J. Ma, *Student performance prediction with regression approach and data generation*, Appl. Sci., *14*(3) (2024), Art. 1148.
[24] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, *Educational data mining techniques for student performance prediction: Method review and comparison analysis*, Front. Psychol., *12* (2021), Art. 698490.
[25] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, J. R. Stat. Soc. Ser. B Stat. Methodol., *67*(2) (2005), 301–320.