Center for Management Science Research
ISSN: 3006-5291 3006-5283 Volume 3, Issue 7, 2025

PREDICTING STUDENT ACADEMIC PERFORMANCE: A STUDY OF HYBRID AND ENSEMBLE REGRESSION MODELS ON HABIT-BASED

Muhammad Asad Ullah[1], Muazzam Ali[*2], Laviza Fatima[3], Amna Ashraf[4], M U Hashmi[5]

[1,3,4]*Department of Basic Sciences, Superior University, Lahore, Pakistan*
[*2,5]*Department of Computer Science, Superior University, Lahore, Pakistan*

[*2]muazzamali@superior.edu.pk

DOI: https://doi.org/10.5281/zenodo.17785471

Keywords
Student performance prediction, machine learning, regression models, ensemble methods, gradient boosting, hybrid models, academic data mining

Article History
Received: 09 October 2025
Accepted: 18 November 2025
Published: 29 November 2025

Corresponding Author: *
Muazzam Ali

Abstract
This work examines the extent in which a variety of regression models can be used to predict the academic performance of students based on an enhanced data set of academic habits. The eight models are: classical linear regression, Support Vector Regression (SVR), and ensemble methods (Random Forest, Bagging, and Increasing), as well as the hybrid models combining any two or more of the basic ones. Every model is tested according to a few performance measures with special focus on the overall accuracy and explanatory power. It has been found that the Among the models tested, the Boosting algorithm outperforms others, achieving the best results with an MAE of 3.1947, MSE of 16.8005, RMSE of 4.0988, and an R^2 of 0.8664. The addition of disparate strengths usually looks appealing in theory since it is possible to see how the skills of each model can be combined in ways that offer sizable performance improvements, but can perform below component models when there is no tuning in the composition of diverse strengths. Therefore, linear regression still serves as quite solid basis, which, at once, manifests evidently the linear elements hidden within the data. Such results raise a concern of careful choice of models and care in controlling complexity when implementing machine learning in education. Also, the combination of hybrid architectures and adaptive weighting schemes could be explored in case one would like to achieve both robustness and interpretability in the process of academic performance prediction.

1.0 Introduction

Due to the digital learning tools booming, schools gather massive repositories of behavioral data, clicks, comments, screenshots of sessions to mention a few. These data offer gold on involvement, performance and end results. The problem is that traditional statistical instruments tend to fail, because they are based on linearity and independent characteristics, which are hardly ever given in the actual classroom presence. This is the reason why researchers have resorted into use of machine learning (ML) that is capable in handling non-linear trends and high-dimensional inputs without losing a wink [1]. ML model with flexible fitting capabilities can pick up these subtle signals which stodgy stats routines may not capture. In this study, a detailed observational data which is filled with daily routines, using study-time, how motivated, and sleep schedule. Some types of regression-based ML algorithm: the linear regression, Support Vector Regression (SVR), and ensemble tools, including Random Forest and Gradient Boosting [2]. Here is the fundamental question: Do hybrid models--techniques that combine many algorithms together--outperform standalone techniques at predicting final grades?

The grade data do not reveal everything; they do not reflect the whole picture of what happens with students in the educational process, as all of this depends on emotional, psychological, and situational factors as well. Most existing models rely on demographics or raw grade points, overlook the better behavioral data recorded currently in Learning Management Systems (LMS) and by student surveys [3]. These systems monitor the time span of attention, self-control and the frequency at which a learner clicks on some content. The hitch of it is that this typically finds one subjected to noise, strong correlation and nonlinearity. The issues are addressed in ensemble models and SVR through searching of hidden signals and also making of robust predictions when faced by messy situations. Random Forest integrates a setting of several decision trees together after which their outcomes are averaged which restricts overfitting and enhances out-of-sample effectiveness. Boosting methods involve addition of layers, rectifying previous errors and refining total fit more. Combination strategies provide a viable trade: the linear stability of models together with the nonlinear pliability of learners. The importance of that hybridity is that behavior of students is usually complex and not linear.

Earlier efforts reported in the education data mining domain point to inconclusive answers as to what kind of ML technique claims the throne, and usually the response has to do with feature choice, the sample, and performance metrics. Indeed, this paper will consider the major performance metrics, i.e., Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R, as the final scoreboard [4 ,5].Good predictions rely very much on clean data preparation. Carefully designed features always have a better chance than features thrown together in a willy-nilly basis. The reliability through K-Fold cross-validation that helps visualize the way models perform among different groups of students. This success of behavior-based ML models in healthcare identification and training of work forces proves their flexibility.

2.0 Literature Review

In the recent past, hybrid and envelope regression models are gaining a lot of popularity in the forecasting of academic performance. The reason behind that buzz is that they can increase learning performance using smart data analytic [6]. Ensemble algorithms, which combine advantages of multiple learning methods, have particularly been hailed as improving accuracy and strength of prediction [7].Such algorithms typically combine demographic, academic and behavioral data in order to identify those students who may be in danger of low attainment. This early identification allows teachers to initiate early and specific remedial action that may guide learners into successful roads. In addition to that, hybrid models, which combine statistical analysis methods such as factor analysis and machine learning, further enhance the process of predicting by focusing on the most indicative predictors of performance [8].Various studies have also pointed out how ensemble models such as the Random Forest and Gradient boosting are excellent at capturing complex educational patterns [9]. They tend to outperform traditional models that are out of range like logistic regression or simple decision trees. With the Open University Learning Analytics Dataset even, stacking ensembles emerge as amazingly good at bordering and multi-way classification. When researchers stack the big picture in the education of making the best prediction models, ensemble learners consistently report greater overall precision, often achieving above 85 percent, than Support Vector Machines (SVM) or k-Nearest Neighbors (k-NN). These results remind us that merging approaches can help pull out subtle trends providing associations between student traits and academic achievement. Hybrid strategies, which connect numbers analysis with forecasting, add still another level of potential. These techniques enhance forecasting by integrating factor analysis and cross-validation and various machine-learning periods. In many datasets of different institutions, models combining two or more techniques perform better than single-technique models in forecasting student achievements [10,11].Even important causal factors correlated with student achievements such as attendance, prior grades, socioeconomic status, and engagement measures can be identified using these methods. An awareness of these factors would allow educators to establish personalized interventions and supports that can be consistent with individual needs. That orientation to personalized directions indicates the sustainable development goals related to all-inclusive and non-discriminative education. As the educational data mining continues to grow, unified procedures and reporting are being put in the limelight. The discoveries do not have the generality that it needs, due to lack of well-documented procedures with accompanying validations. The unavailability of standardized available data and benchmarks of the tests also remains an obstacle to compare machines and human thoroughly. Nonetheless, the association of machine learning with the field of educational research will keep adding opportunities to student-support systems. Educational institutions can rely on the data-based insights to reduce the dropout rates, increase the quality of learning, and personalize education paths to the individual. All in all, the evidence provided in the literature demonstrates that hybrid and ensemble models are the crucial component of predictive education research. Researchers in the future need to continue the process of improving these methods making them ethical, transparent, and effective on a variety of students.

3.0 Methodology

3.1 Data Preprocessing

Considering the dataset used in the present study, one stalks before a very diverse sets of variables, i.e., behavioral, demographic, academic ones along with the target variable: the exam scores of the students. In order to defend the intellectual quality of data and establish a strong model, implemented extensive preprocessing procedures on all modes. First, we found the outliers in numeric features and eliminated them by the *z*-score technique. The *z*-score was calculated of each of these columns (xi) of the numerical numbers as:

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \quad (1)$$

in which the μ_i and σ_i represents mean and standard deviation of the feature respectively. Any sample exceeding |zi|>3 was considered as an outlier and removed to reduce noise and restrict impact of high values during training. Then, applied the label encoding on all categorical elements and converted them to integer labels, which basically means giving each category in the column an integer label in a unique fashion and enabling the features to be converted into numeric quantities which are what machine learning algorithms need [12].

Then scaled our features with StandardScaler to make each column in the dataset have zero-mean and unit-variance:

$$X_{scaled} = \frac{X - X^{-}}{\sigma_x} \quad (2)$$

Where X^{-} and $\sigma_x$denotes the mean and standard deviation of each feature wherein represents the feature, and represents the mean and standard deviation of each feature, respectively. This normalization is critical to the extent that all features are equipproportional to the learning of a model and particularly significant in the case of algorithms that are magnitude-sensitive (Support Vector Regression (SVR) in particular).

Lastly, Divide the processed dataset into training and the testing datasets (70:30 ratio) using stratified random sampling to maintain the distribution of the target variable across the training and the testing snapshots.

3.2 Model Descriptions and Training Procedures

The tree-based ensembles, hybrid ensembles, kernel-based approaches, and the traditional linear models were offered on the menu. To make things apples-to-apples, the training on all models was conducted with the same batch of preprocessed data.

3.2.1 Support Vector Regression (SVR)

Support Vector Machines' concepts are used to regression issues via SVR. Formulated as follows, it looks for a function f(x) that maintains model flatness while deviating from real goal values by no more than ϵ:

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{subject to} \quad |y_i - (\boldsymbol{w}.\boldsymbol{x}_i + b)| \le \epsilon \; \forall_i \quad (3)$$

Slack variables and a penalty parameter, C, to determine the amount of additional leeway awarded to the model and the extent to which we require the trade-off between simplicity and accepting mispredictions to be stringent.

Non-linear cases can be supported with kernel as radial basis function (RBF)Radial Basis Function (RBF) Kernel [14], where the functions are defined as:

$$K(x_i - x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \quad (4)$$

After kernel, the training often takes place under stratified k-fold cross-validation with the idea of ensuring that the end product is stable and well able to generalize.

3.2.2 Decision Tree Regressor

Decision Trees hour cut data in a hierarchical method with the help of feature limits to decrease the variance within every new subgroup [15,16]. The rule of splitting that is selected is intended to maximize:

$$\min_{split} \sum_{K=1}^{K} \sum_{x_i \epsilon R_K} (y_i - \bar{y} \ R_K)^2 \quad (5)$$

whereR_K are the resulting set of regions and $\bar{y} \ R_K$ is the average target value of a region. Last put the limit on the tree depth to avoid the overfitting.

3.2.3 Linear Regression

Ordinary Least Squares, because it is one of the usual approaches when dealing with least squares to linear fit linear relationship between a set of predictors and a target variable by minimizing the residual sum of squares in the set of coefficients.

$$y = \mathbf{X}\beta + \epsilon \quad (6)$$

The linear regression is an in-box fitting of this model and can provide the baseline to compare.

3.3 Ensemble Models: Bagging, Boosting, Random Forest, and Stacking

Ensemble learning leverages multiple models to improve prediction performance.

3.3.1 Bagging Regressor

It factors many decision trees on bootstrapped training sets and averages the predictions. The bare notion is:

$$\hat{f}\,bagging(x) = \frac{1}{M} \sum_{m=1}^{M} f_m \ (\boldsymbol{X}) \quad (7)$$

M is the base learners. The bagging technique reduces variance and overfitting.

3.3.2 Gradient Boosting Regressor

Imagine a set of decision trees stacked on top of each other where the final one is trained on all the mistakes of the previous one. Both of the trees are attempting to reduce the expense of training by minimizing a loss upon a loss function L, the most frequent of which is the squared error, through gradient descent in the arena of functions:

$$f_m(x) = f_{m-1}(x) + v h_m(x) \quad (8)$$

Where $h_m(x)$ is the new tree fit to the negative gradient of the loss, and ν is the learning rate.

3.3.3 Random Forest Regressor

A mix between bagging and random feature selection at decision nodes. In this way not only make the decorrelation between individual trees better, but also increase the overall predictive accuracy [17].

3.3.4 Stacking Regressor

Stacking is a form of combination of various base learners in which a meta-model is constructed to be trained using the predictions of the other base learners [18].

3.3.5 Hybrid Model Construction

Actually, hybrid models are just mix and match ensembles composed of the output of a couple of various classifiers such as stacking, bagging and boosting where each of them overcomes the shortcomings of the other one [19]. Their weights are usually tuned empirically in order to optimise performance measures.

$$\hat{y}\,hybrid = \sum_{i=1}^{N} \omega_i \hat{y}_i \quad (9)$$

weights ω_i chosen empirically to maximize performance metrics

3.4 Cross-Validation and Stratification

Since the dataset dealing includes continuous target variable, discretize that target variable into discrete categories. In particular, Perform the variable split into bins according to the quantiles of distribution, implying that each of the folds has the focal points that follow the general distribution of scores approximately [20]. This allowed doing so in such a way that all folds reflected the real distribution and increased the training and evaluation stability. Employed 5-fold splits- after an initial training/test on a single fold, shrug off that fold and repeat the process, four more times. The approach reduces the variability of performance estimates due to mean-pooling of folds. Finally, created one prediction ensemble by averaging all predictions per each of the folds. It is essential to do that since the estimations of every fold can vary significantly, particularly when there are extreme values in some of them. By averaging, these irregularities are evened out and only one, reliable prediction.

3.5 Evaluation Metrics

Model performance was assessed using several standard regression metrics:

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (10)$$

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (11)$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (13)$$

$$Adj.R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \quad (14)$$

In approaching a prediction problem, two parameters should be in mind the number of predictors, the p, and the size of the training set, n. There are too many features that may bloat R^2 and increases the appearance of fit of the model, hence p/n is helpful in checking that whether the level of complexity used is justified.

Four measures, namely, (1) mean absolute error (MAE), (2) root mean squared error (RMSE), (3) usual R-squared (R2), and (4) adjusted R-squared (adj R2) were used to check the performance. The MAE and RMSE define the extent to which the error terms are big whereas R^2 and adj R^2 will tell how richly the model accounts data and how specifically it fits without being overfitted. The preparation data was done with caution. Imputation of missing values and standardization of features were followed by division of dataset into training and test sets. The training was performed using stratified k-fold cross-validation in order that the performance of each model would be compared with the same proportion of observations. A wide set of regression methods has been trained, beginning with linear models and continuing with more complex ensemble and mixed methods. This spectrum gave room to an objective comparison of model performance.

Each submission was evaluated on the entire range of metrics: raw MAE/RMSE, R^2 and adj R^2 along with corresponding ones of the test set so as to reflect out-of-sample performance. The best result in terms of MAE, RMSE and test MAE/RMSE in each of the methods was measured, as well as the highest R^2 and adj R^2. Combined models were also explored where hybrid models were constructed by enhancing the capabilities of various algorithms. All the pairs were tested against the same criteria as individual models to isolate the improvements on accuracy and explainability. Conclusively, through intensive data cleaning, the selection of reasonable p/n ratio and evaluation across several parameters, the study offered an insightful review of training student behaviors to machine learning methods.

Table 1: Parameters of the Models

Model	Hyperparameter	Value / Description	Notes
Support Vector Regression (SVR)	kernel	'rbf' (in some codes), default 'rbf' if not set	RBF kernel used in codes 3,7,8; default in code 1,2
	C	1.0 (code 7,8), default in others	Regularization parameter
	epsilon	0.2 (code 7,8), default otherwise	Insensitive tube width
	Other	Default	In code 1,2 SVR() called without params
Decision Tree Regressor	random_state	42 (code 2)	For reproducibility
	max_depth	5 (code 3), default otherwise	Limits tree depth
	Other	Default	No other params set
Linear Regression	None	Default	No hyperparameters
RidgeCV (Final Estimator in Stacking)	alphas	Default (automatically chosen)	Used in code 1 stacking regressor
	cv	Default	Internal cross-validation
Bagging Regressor	estimator	DecisionTreeRegressor()	Base learner for bagging (code 1,4)
	n_estimators	50	Number of base estimators
	random_state	42	Reproducibility
Gradient Boosting Regressor	n_estimators	100	Number of boosting stages
	learning_rate	0.1 (code 5)	Step size shrinkage
	max_depth	3 (code 5)	Depth of individual trees
	random_state	42	Reproducibility

Stacking Regressor	estimators	[('svr', SVR()), ('dt', DecisionTreeRegressor())] (code 1)	Base learners
	final_estimator	RidgeCV (code 1), LinearRegression (code 3)	Meta-model
	cv	3 (code 3)	Cross-validation folds for stacking
Random Forest Regressor	random_state	42	Reproducibility
	Other	Defaults	Number of trees, max_features default
Train-test split	test_size	0.3	30% test data
	random_state	42	Seed for splitting
StratifiedKFold	n_splits	5	Number of folds
	shuffle	True (some codes)	Shuffle before split (codes 3–9)
	random_state	42 (some codes)	For reproducibility

4.0 Results and Discussions

The second section considers some of the regression models that have been employed in predicting the student's exam scores based on a comparison of habits and behavioral information. The performance of each of the model is measured with more than one metric and thus a comprehensive comparison is possible. It performs the analysis, noting the differences in accuracy and explanatory power, which informs the way the various modeling strategies work on the data.

Table 2: Performance Evaluation of Models

Model	MAE	MSE	RMSE	R^2	Adjusted R^2
Random Forest	3.2559	17.1210	4.1377	0.8639	0.8637
Support Vector Regression (SVR)	3.2900	19.6491	4.4327	0.8438	0.8436
Linear Regression	3.1761	17.0133	4.1247	0.8647	0.8645
Bagging	3.2543	17.1153	4.1371	0.8639	0.8637
Boosting	3.1947	16.8005	4.0988	0.8664	0.8662
Hybrid (SVR + Decision Tree + Linear)	3.3196	19.0855	4.3687	0.8482	0.8480
Hybrid (Boosting + Bagging + Stacking)	3.2337	17.2217	4.1499	0.8631	0.8629

A simple, similar measure of the extent to which model predictions and the actual score on the exam are close is the mean absolute error (MAE). Boosting regressor has the lowest MAE of 3.1947, demonstrating its capabilities to repeatedly make the predictions that remain near the familiar information. Linear Regression is not a loser either, and its MAE is 3.1761, so its simplicity did not come at the cost of accuracy. The model containing a combination of Support Vector Regression, Decision Tree Regressor and the Linear Regression has yielded MAE of 3.3196, which implies that simply averaging the predictions made by different models without specific weighting is ultimately introducing noise and can hurt accuracy. Random Forest and Bagging regressor produce similar scores, a little bit higher than the Boosting, yet showing a decent predictive dependability.

Looking at the figures of Mean Squared Error (MSE) attached to each regression model, it is obvious that they do not appear in the same sequence in rankings. Gradient Boosting won the day with a low MSE of 16.8005 ensuring that it truly is possible to athletically and better fit the predicted scores to the actual examination results in this data set in the process of reducing the error one step at time which is what boosting is all about.

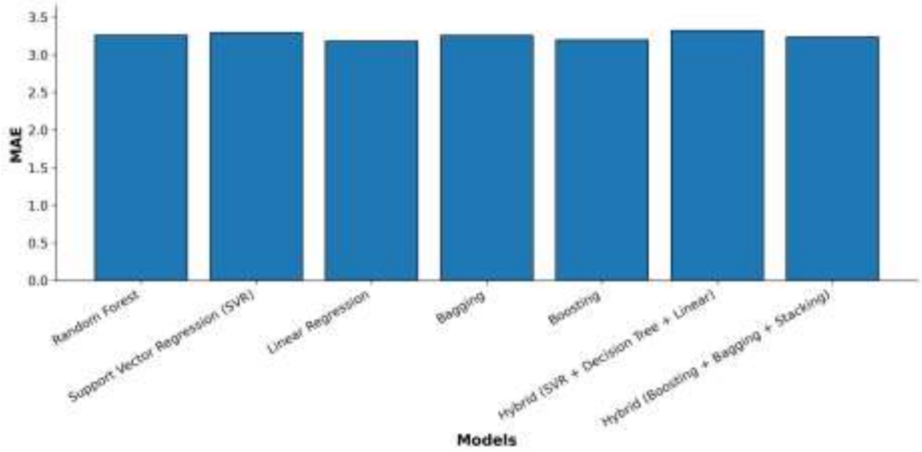

Figure 1: Performance Metric MAE of Models

The situation is not so bright with the ensemble model which combines Support Vector Regression, Decision Tree Regressor, and Linear Regression. That hybrid one was on the opposite side of the scale with 19.0855 MSE. What this basically means is that, this combined model has seemed to lose more accuracy than it acquired in the combination of these three learners. The Random Forest, Bagging, and Linear Regression are clustering together on the condition that each of them addresses the data adequately but cannot rival the sleekness of the Boosting technique.

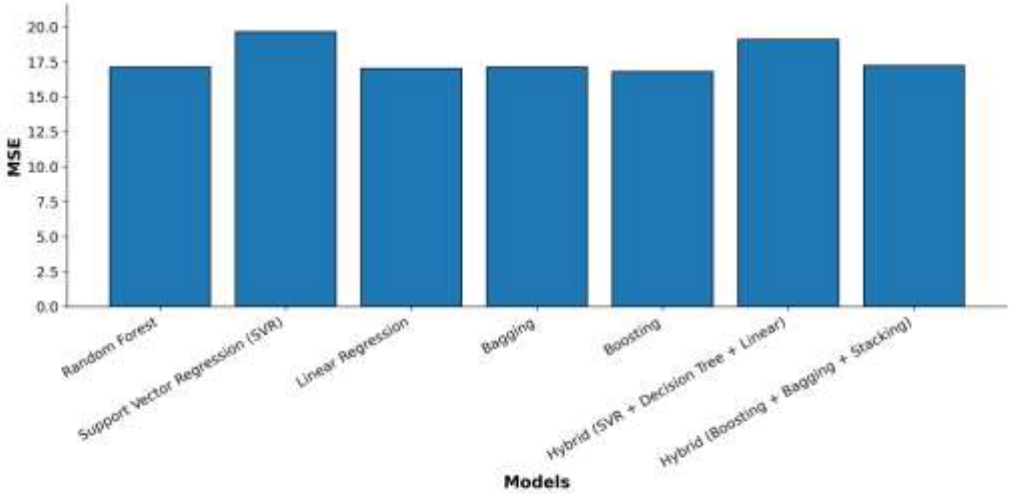

Figure 2: Performance Metric MSE of Models

Again, when look under the surface at the root mean squared error (RMSE), which gives a measure of error in the original units of exam scores, find the same picture as that presented by the R-squared values. Boosting can also be distinguished since it has the lowest RMSE value of 4.0988 which further confirms its high reliability in giving predictions which are more accurate. Linear Regression fares a bit better than Random Forest and Bagging: its RMSE is 4.1247, compared to those of the latter two that are a bit higher than 4.13 as well, suggesting that the relationship we see in the data is also strongly linear, and as such, is what linear models excel at predicting. SVR, Decision Tree, and Linear Regression approaches have a higher RMSE of 4.3687 indicating less accurate predictions of all. So, the middle ground between those two extremes occupied by the Boosting, Bagging, and Stacking hybrid displays a degree of synergy but without being better than standalone boosting.

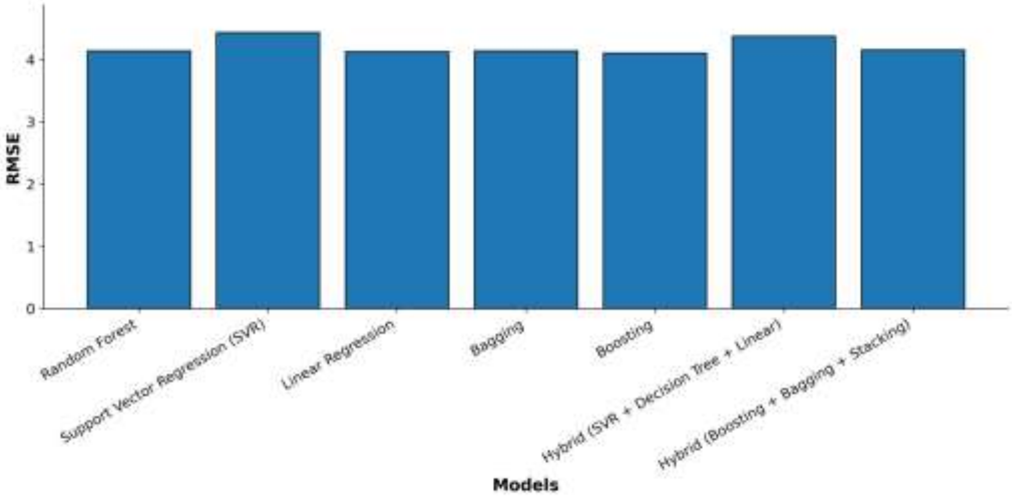

Figure 3: Performance Metric RMSE of Models

By having coefficient of determination (R^2) and its adjusted form, one gets an idea about the strength of the models in explaining variability in the scores across exams as well as adjusting these explanations to the complexity of the models. Increasing leads with the maximum R^2 value of 0.8664 and adjusted R^2 of 0.8662 show the explanation capacity of the leads is very high and it fits well. Linear Regression comes closely with R 2 of 0.8647 and adjusted R^2 of 0.8645 further proves that there are linear trends that exist in the data. Random Forest and Bagging has a similar value of 0.864, which provides an impression that ensemble tree-based methods are reliable when it comes to capturing variance. The model that includes both SVR and Decision Tree shows the lowest R^2 (0.8482), adjusted R^2 (0.8480) which could be attributed to poor combination of these two models and contradicting bias of the models. The Boosting, Bagging, and Stacking hybrid has moderate explanatory power that seems to be in the middle between the single ensembles of models and the inferior SVR-based hybrid.

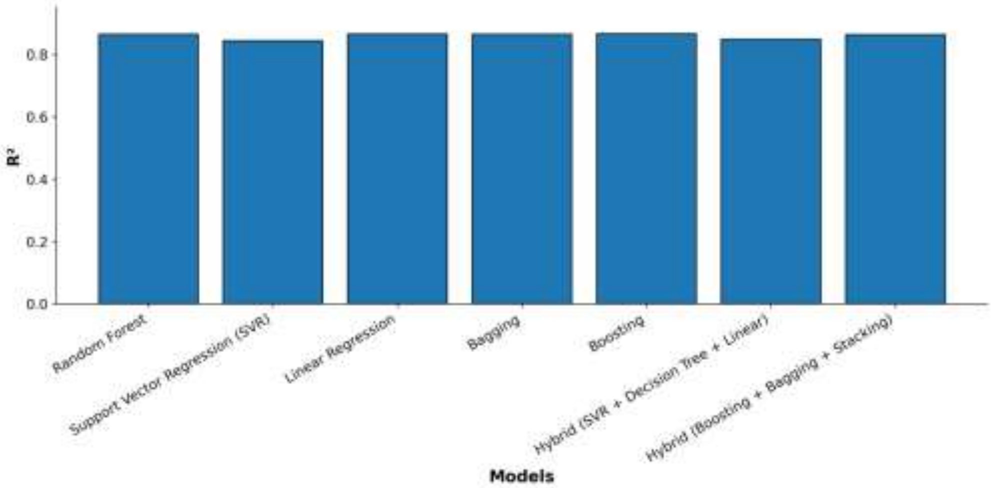

Figure : Performance Metric R^2 of Models

Adjusted R^2 having a model complexity penalty to avoid overfitting, ressembles the R^2 results with Boosting scoring the best adjusted at 0.8662. This shows that it would use better fit without sacrificing unnecessary complexity. Adjusted R^2 in Linear Regression is also high at 0.8645, as well, indicating it to be parsimonious and effective in its modeling. Adjusted R^2 is a bit less in Random Forest and Bagging with 0.8637. The hybrid SVR-based model once more indicates lowest adjusted R^2 of 0.8480 implying that complexity provided by mixing different algorithms rather than providing a perfect blend actually influences the model in the adverse direction rather than positively improving its generalizability.

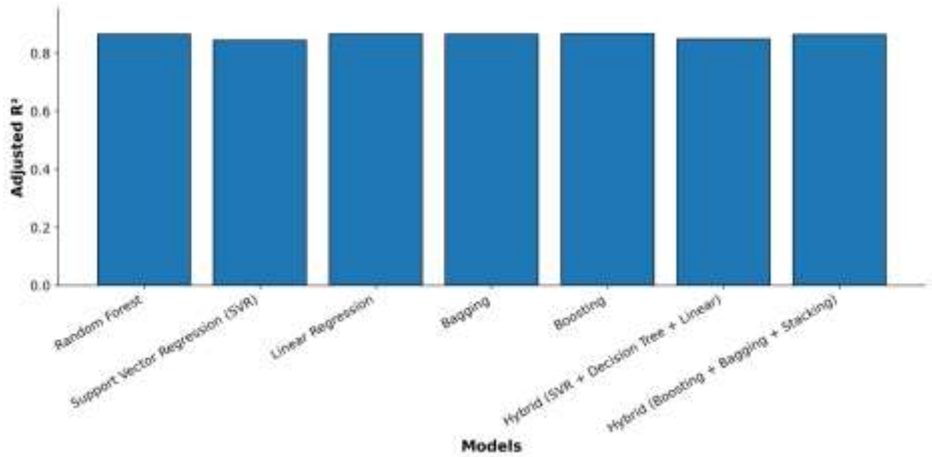

Figure 5: Performance Metric Adjusted R^2 of Models

On the whole, these experiments demonstrate: whereas ensemble models are better than simpler ones, attentiveness of gradient-boosting in correcting errors enables its striking the best possible balance between accuracy and variance explanation. Hybrid models require a hyper-tuning to utilize strengths of constituent modules correctly because an insufficient combination may diminish performance. High value of the linear regression means that the associations in the dataset are rather linear, and therefore simpler variants of the model should be taken into consideration among complex ensembles. The research done in the future could focus on enhancing the hybrid methods and ensemble weighting to make their predictions more accurate and to make models more resilient.

5.0 Conclusion

This paper demonstrates that the most promising method of predicting student academic performance using complex habit-derived data is by employing the ensemble of methods and, in particular, gradient boosting. On the strengths of accuracy and explanation of variance, boosting is better than single models and combination of hybrid because of its sequential resolution of errors. Nonetheless, hybrid models may perform quite well when designed and properly tuned to avoid failure to perform. The strength of linear regression implies that the simple models are also practical, especially in cases of linearity in data relationships. These findings can be useful to data scientists and teachers who want to use predictable tools during targeted academic interventions. To make hybrid approaches in use in a diverse range of educational scenarios more applicable, future studies are to focus on the development of model interpretability.

Acknowledgment

The authors acknowledge their different departments and education institutions that gave this research anonymised student data. We also simply like the constructive criticism gave by the reviewers and it gave our work quality in large measure.

Conflict of Interest

In this study, the authors do not state conflict of interest. All the analyses were done in an objective manner, and the other external factors did not influence the study design, data interpretation, and the reporting of the results.

References

[1] Kapoor A. ML Approach: Algorithms, Real-World Applications and Research Directions. Real-World Applications and Research Directions (November 01, 2024). 2024 Nov 1.

[2] Kumar S, Srivastava M, Prakash V. Enhancing Mutual Fund Price Prediction: A Hybrid Ensemble Approach with Random Forest, SVR, Ridge, and Gradient Boosting Regressors. InInternational Conference on Signal, Machines, Automation, and Algorithm 2024 (pp. 551-566). Springer, Singapore.

[3] Levin S. Identifying Student Behaviour Patterns Based on LMS Data.

[4] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science. 2021 Jul 5;7:e623.

[5] Dessain J. Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric. Expert Systems with Applications. 2022 Aug 1;199:116970.

[6] SULEIMAN IB. *PREDICTING STUDENTS'ACADEMIC PERFORMANCE USING LINEAR REGRESSION* (Doctoral dissertation, OPEN UNIVERSITY).

[7] Farhadi S, Tatullo S, Ferrian F. Comparative analysis of ensemble learning techniques for enhanced fatigue life prediction. Scientific Reports. 2025 Apr 1;15(1):11136.

[8] Hao D, Xiaoqi Y, Taoyu Q. Hybrid Machine Learning Models Based on CATBoost Classifier for Assessing Students' Academic Performance. International Journal of Advanced Computer Science & Applications. 2024 Jul 1;15(7).

[9] Tong T, Li Z. Predicting learning achievement using ensemble learning with result explanation. PloS one. 2025 Jan 2;20(1):e0312124.

[10] Munemma D, Maheswari KU. FEATURE SELECTION ALGORITHMS FOR PREDICTING STUDENTS ACADEMIC PERFORMANCE. MATERIAL SCIENCE. 2024 Jan;23(01).

[11] Dioscord NM. *Predictive Model for Early Detection of Higher Education Dropout Using Machine Learning Techniques* (Doctoral dissertation, University of Rwanda).

[12] Bolikulov F, Nasimov R, Rashidov A, Akhmedov F, Young-Im C. Effective methods of categorical data encoding for artificial intelligence algorithms. Mathematics. 2024;12(16):2553.

[13] Yadav A, Ahmed M. Predictive Modeling for Thread Optimization in OpenMP-Based Parallelization Using Machine Learning. In2024 IEEE 17th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC) 2024 Dec 16 (pp. 339-344). IEEE.

[14] Ding X, Liu J, Yang F, Cao J. Random radial basis function kernel-based support vector machine. Journal of the Franklin Institute. 2021 Dec 1;358(18):10121-40.

[15] Bhaduri D, Toth D, Holan SH. A Review of Tree-Based Methods for Analyzing Survey Data. Wiley Interdisciplinary Reviews: Computational Statistics. 2025 Mar;17(1):e70010.

[16] Sushma PG, Sriramakrishnan GV. Exploring predictive algorithms: Linear regression and decision tree in student performance. InAIP Conference Proceedings 2025 Apr 11 (Vol. 3270, No. 1, p. 020154). AIP Publishing LLC.

[17] Beatty J, Sundkvist S. Comparative Analysis of Tree-Based Methods for Predictive Modeling.

[18] Azad RU, Hossain MS, Talukdar M, Nahar N, Hanip A. A Stacked Meta-Model Framework for Diabetes Prediction: From Effective Feature Engineering to Meta-Learning. In2024 27th International Conference on Computer and Information Technology (ICCIT) 2024 Dec 20 (pp. 3098-3103). IEEE.

[19] Shaikh TA, Rasool T, Verma P, Mir WA. A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art. Annals of Operations Research. 2024 Dec 24:1-77.

[20] Awedh M, Mueen AH. Early Identification of Vulnerable Students with Machine Learning Algorithms. WSEAS Trans. Inf. Sci. Appl. 2025; 22:166-88.