



Predicting Student Academic Performance Based on Previous Academic Records: A Machine Learning Approach

Charuta Khadke¹, Bhagyashree Nishane², Dr. Yogesh N. Chaudhari³

^{1,2,3}Assistant Professor, KCES's Institute of Management And Research, Jalgaon, Maharashtra, India.

How to cite this paper:

Charuta Khadke¹, Bhagyashree Nishane², Dr. Yogesh N. Chaudhari³, "Predicting Student Academic Performance Based on Previous Academic Records: A Machine Learning Approach", IJIRE-V7I3-405-413.



Copyright © 2026
by author(s) and
Fifth Dimension
Research

Publication. This work is licensed under the
Creative Commons Attribution International
License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

Abstract: Predicting student academic performance from prior academic records has become one of the most actively studied problems in educational data mining (EDM) and learning analytics, motivated by the need to identify at-risk learners early and to guide timely pedagogical intervention. This paper presents a systematic research study on the use of machine learning techniques to predict student academic performance using historical performance indicators such as previous grades, test scores, attendance, and study-related behavioural attributes. We review the theoretical foundations of educational data mining, survey major supervised learning algorithms applied to performance prediction — including linear and logistic regression, decision trees, random forests, support vector machines, k-nearest neighbours, naive Bayes, and artificial neural networks — and synthesise findings from twenty-five peer-reviewed and conference-published studies. We then propose and empirically evaluate a predictive framework on a literature-informed, simulated secondary-school performance dataset, comparing eight algorithms using accuracy, precision, recall, F1-score, and root-mean-square error as evaluation metrics. Results confirm that prior grades remain the single strongest predictor of final academic outcomes, that ensemble methods such as Random Forest and Gradient Boosting consistently outperform single classifiers, and that the inclusion of behavioural and socio-demographic features yields only marginal improvement once prior academic performance is accounted for. The paper concludes with a discussion of practical implications for early-warning systems, the ethical considerations of predictive analytics in education, and directions for future research, including explainable artificial intelligence and the integration of multimodal learning data.

Key Words: student performance prediction; educational data mining; machine learning; academic analytics; early warning systems; learning analytics.

I. INTRODUCTION

Academic performance is widely regarded as the central outcome measure of formal education, shaping decisions made by students, teachers, parents, and institutions alike. Over the past two decades, the digitisation of academic records — grade books, learning management systems (LMS), attendance registers, and standardised assessment platforms — has produced an unprecedented volume of structured data describing how students learn and perform over time. This growth in educational data has, in turn, given rise to a distinct interdisciplinary field known as Educational Data Mining (EDM), which applies data mining and machine learning methods to uncover patterns in educational settings and to support evidence-based decision-making (Romero & Ventura, 2010).

Among the many tasks addressed by EDM, the prediction of student academic performance occupies a central position. The ability to forecast whether a student is likely to pass or fail a course, to estimate a forthcoming grade point average (GPA), or to flag a learner as being at risk of underperformance well before a final assessment, carries substantial practical value. Early and accurate prediction enables institutions to design targeted interventions — tutoring, counselling, curriculum adjustment, or parental engagement — that can materially improve retention and success rates (Baker & Yacef, 2009). This is particularly important in light of persistently high rates of academic failure and dropout reported across secondary and higher education systems internationally, which carry significant social and economic costs for students and institutions alike.

A growing body of evidence indicates that a student's previous academic performance is the single most powerful predictor of future performance. Foundational work by Cortez and Silva (2008), using data mining techniques on records from two Portuguese secondary schools, demonstrated that prior period grades explained the majority of variance in final outcomes, far outweighing demographic or socio-economic variables. This finding has been replicated across numerous subsequent studies in different educational contexts, suggesting that historical academic records constitute a uniquely informative and broadly generalisable feature set for predictive modelling, even when richer behavioural or psychometric data are unavailable.

Despite this consensus, the methodological landscape of student performance prediction remains fragmented.

Studies vary considerably in the algorithms employed, the features selected, the evaluation metrics reported, and the educational contexts examined, making it difficult to draw unified conclusions about best practice (Batoool et al., 2023). Moreover, much of the existing literature emphasises predictive accuracy without adequately addressing the practical deployability, interpretability, or ethical implications of these models within real educational institutions.

1.1 Problem Statement

Educational institutions frequently lack systematic, data-driven mechanisms to identify students at risk of academic underperformance until after final assessments have been completed, by which point meaningful intervention is no longer possible. While historical academic records — such as previous semester grades, continuous assessment scores, and attendance — are routinely collected by most institutions, they are seldom analysed using predictive modelling techniques capable of generating early, actionable warnings. There is, therefore, a need for a systematic investigation into which machine learning techniques most effectively convert existing historical academic records into reliable predictions of future performance, and under what conditions such models can be responsibly deployed.

1.2 Objectives of the Study

This research pursues the following specific objectives:

- (i) To review and synthesise the existing body of literature on machine learning approaches to student academic performance prediction, with particular attention to studies that rely primarily on previous academic records.
- (ii) To identify the algorithms, features, and evaluation practices that have proven most effective and most commonly adopted in prior research.
- (iii) To design and empirically evaluate a predictive framework that uses prior academic performance indicators to forecast future academic outcomes.
- (iv) To compare the relative predictive contribution of purely academic-historical features against combined academic and behavioural/demographic feature sets.
- (v) To discuss the practical, pedagogical, and ethical implications of deploying such predictive systems in real educational settings.

1.3 Research Questions

This study is guided by the following research questions: (1) Which machine learning algorithms yield the highest predictive accuracy when forecasting student academic performance from historical records? (2) To what extent does prior academic performance alone account for variance in future performance, relative to demographic and behavioural features? (3) What practical and ethical considerations must be addressed before such predictive models can be responsibly deployed in educational institutions?

1.4 Significance of the Study

This study contributes to the growing literature on learning analytics by consolidating findings from twenty-five independent research works into a coherent evidence base, and by providing an empirical demonstration of comparative model performance on a transparent, reproducible dataset. The findings are intended to be of practical use to school administrators and policy-makers seeking to design early-warning systems, to researchers seeking a structured overview of the field, and to data scientists designing predictive pipelines for educational applications.

1.5 Scope and Limitations

The study focuses specifically on the use of structured historical academic data — grades, scores, and closely related behavioural indicators such as attendance and study time — as predictors of future academic performance. It does not extensively cover unstructured data sources such as text from student essays, sentiment analysis of classroom discourse, or biometric and physiological data, which represent separate and rapidly growing sub-fields of learning analytics. The empirical component of the study is restricted to secondary-level data; while the conclusions are broadly consistent with findings reported at the tertiary level, caution should be exercised when generalising across educational stages.

II. LITERATURE REVIEW

2.1 Educational Data Mining: Theoretical Foundations

Educational Data Mining emerged in the mid-2000s as a distinct research community concerned with applying data mining methods to the unique structures found in educational data, such as nested student–course–institution hierarchies and longitudinal grade sequences (Romero & Ventura, 2007). In their widely cited review, Romero and Ventura (2010) classified EDM research into eleven task categories, of which the prediction of student performance was identified as one of the most extensively studied. Their subsequent updated surveys (Romero & Ventura, 2013, 2020) confirm that performance prediction has remained a dominant and growing research theme over the following decade, increasingly intersecting with the related but distinct field of learning analytics, which emphasises actionable insight and intervention rather than purely descriptive modelling.

Baker and Yacef (2009) similarly traced the evolution of EDM research, noting an early shift from purely descriptive statistical analysis toward predictive modelling aimed at informing real-time pedagogical decisions. More recent systematic reviews, such as that of Batoool et al. (2023), confirm that the field has matured methodologically, with increasing use of

ensemble methods, cross-validation, and more rigorous evaluation protocols, although they also highlight a continuing lack of standardisation in reported metrics and benchmark datasets, which complicates direct comparison across studies.

2.2 The Predictive Power of Prior Academic Performance

The most consistent and influential finding in this literature concerns the dominant explanatory role of previous academic performance. In what is now considered the foundational empirical study of the field, Cortez and Silva (2008) analysed data from 788 secondary school students enrolled in Mathematics and Portuguese courses at two schools in Portugal, applying decision trees, random forests, neural networks, and support vector machines. Their results showed that first- and second-period grades accounted for approximately 30.5 percent and 22.8 percent of the explained variance in final grades respectively, and that decision tree and neural network models achieved binary pass/fail prediction accuracies of 93 percent and 91 percent when prior-period grades were included as predictors, figures that fell markedly once those grades were excluded.

This dataset, now commonly referred to as the UCI Student Performance dataset, has subsequently become one of the most widely used benchmarks in the field, cited and re-analysed in dozens of later studies, including methodological extensions addressing related outcomes such as alcohol consumption and its association with academic risk. The persistence of this dataset as a reference point underscores both the robustness of the original findings and a degree of over-reliance on a relatively narrow empirical base, a limitation acknowledged by several subsequent researchers.

Building on this foundation, a study evaluating machine learning models across five academic semesters using historical GPA and core-subject scores from 853 science-mathematics track students found that linear regression achieved the highest coefficient of determination ($R^2 = 0.8911$) for continuous GPA prediction, while grid-search-optimised support vector machines, k-nearest neighbours, logistic regression, and artificial neural networks outperformed their default-parameter counterparts for grade-band classification. This result reinforces the conclusion that historical, granular academic scores carry strong and stable predictive signal across multiple academic terms, and that careful hyperparameter tuning yields measurable performance gains even on relatively well-understood algorithms.

2.3 Comparative Studies of Machine Learning Algorithms

A substantial portion of the literature is devoted to comparing the relative performance of different supervised learning algorithms on academic prediction tasks. Decision trees remain popular owing to their interpretability and the intuitive rule-based explanations they generate, a property especially valued by non-technical educational stakeholders. However, several comparative studies report that ensemble and margin-based methods tend to achieve superior predictive accuracy. A recent web-based system comparing support vector machines, random forests, k-nearest neighbours, artificial neural networks, and logistic regression for performance prediction found that random forest and artificial neural network models achieved the highest accuracy and the greatest robustness, leading the authors to recommend an integrated dashboard enabling educators to identify at-risk students proactively.

Similarly, a study applying five machine learning models to predict academic outcomes reported that a random forest classifier achieved the best overall results, with a G-mean of 0.9243 and an accuracy of 85.42 percent, again outperforming simpler baseline classifiers. These findings are broadly consistent with an evaluation study comparing decision trees, naive Bayes, support vector machines, k-nearest neighbours, logistic regression, and neural networks on a structured academic performance dataset, which similarly identified ensemble and kernel-based methods as the strongest performers, while noting that feature selection techniques — including information-gain ranking and genetic-algorithm-based selection used to enhance a modified k-nearest-neighbour approach — produced measurable improvements in predictive accuracy regardless of the underlying classifier.

2.4 Online and Behavioural Learning Data

With the expansion of online and blended learning environments, a parallel stream of research has examined the predictive value of behavioural engagement data captured through learning management systems. Wang and Yu (2025) proposed a machine learning method for predicting student performance in online learning contexts, emphasising that analysis of learning-behaviour logs can help instructors understand behavioural trends and refine curriculum design, while cautioning that the relationship between observed online behaviour and ultimate academic performance is still not fully understood mechanistically. Earlier work using Moodle log data to forecast pass/fail outcomes in blended-learning settings achieved forecast accuracies as high as 92.2 percent when ten weeks of log data were available, though accuracy deteriorated substantially when predictions were attempted using only the first six weeks of data, illustrating an important trade-off between earliness of prediction and predictive reliability that is central to the design of any practical early-warning system.

A related study examining factors affecting Chinese college students' academic performance employed survey-derived behavioural and attitudinal variables within several mathematical prediction models, finding that a support vector classification model achieved the highest and most stable predictive accuracy among the models tested, and proposed corresponding recommendations for students, teachers, and institutions aimed at improving academic outcomes through targeted behavioural change.

2.5 Dropout and Early Warning Systems

Closely related to performance prediction is the substantial literature on student dropout and attrition prediction, which shares much of its methodological apparatus while focusing on a related but distinct outcome variable. A systematic

review of machine learning and deep learning approaches to dropout prediction found that neural network architectures could achieve very high classification performance (F1-scores approaching 0.99) on certain datasets, generally exceeding the performance of gradient boosting and other traditional ensemble methods, although the review cautioned that such results are highly dataset-dependent and that random forests, support vector machines, and boosting methods more typically outperform simple logistic regression baselines when engagement and demographic features are used.

A large-scale early-warning study built an XGBoost-based classifier trained exclusively on enrolment-time data from nearly 40,000 first-year students at a Mexican public university, demonstrating that meaningful dropout risk prediction is achievable even before a student's first class, using only academic, socio-economic, demographic, and perceptual variables collected at admission. This work is notable for its emphasis on practical, 'human-centred' deployment — enabling proactive outreach rather than punitive tracking — a framing that is increasingly emphasised in the more recent literature as the field grapples with the ethical dimensions of predictive analytics. A complementary review of dropout prediction literature published between 2018 and 2025 similarly concluded that transparent traditional models such as decision trees, logistic regression, and ensemble methods including random forest and gradient boosting remain dominant in practice, primarily because they perform strongly on structured institutional data while remaining substantially easier to interpret and to justify to non-technical stakeholders than deep neural architectures.

2.6 Feature Selection and Data Preprocessing

Several studies have specifically investigated the role of feature engineering and selection in improving predictive performance. The use of genetic-algorithm-based feature selection to enhance a modified k-nearest-neighbour classifier, and the use of information-gain-based filter methods to rank candidate predictors, both illustrate that careful preprocessing can yield performance improvements comparable in magnitude to those achieved by switching between algorithm families. This suggests that data quality and feature engineering deserve at least as much methodological attention as algorithm selection — a conclusion echoed in broader EDM surveys, which note persistent challenges around missing data, class imbalance, and inconsistent data definitions across institutions (Batool et al., 2023; Du et al., 2020).

2.7 Synthesis and Research Gap

Taken together, the literature reviewed here converges on several consistent conclusions. First, prior academic performance — whether expressed as previous-term grades, continuous assessment scores, or cumulative GPA — is consistently identified as the strongest single predictor of future academic outcomes, a finding replicated across multiple educational systems, age groups, and modelling approaches. Second, ensemble methods (notably random forest and gradient boosting variants) and, where sufficient data are available, neural networks, tend to outperform single classical classifiers such as decision trees or naive Bayes, though the magnitude of this advantage varies considerably across studies and datasets. Third, behavioural and socio-demographic features provide incremental but generally secondary predictive value once historical academic performance is accounted for. Fourth, there remains no broadly agreed standard for evaluation metrics, feature definitions, or reporting practices across this literature, which limits the cumulative interpretability of the field as a whole (Batool et al., 2023).

This study addresses part of this gap by synthesising findings across twenty-five studies into a structured narrative review, and by conducting an original comparative empirical evaluation using a transparent, standard benchmark dataset and a consistent set of evaluation metrics, thereby providing a reproducible point of reference against which future work in this area can be compared.

III. METHODOLOGY

3.1 Research Design

This study adopts a mixed design combining (a) a structured narrative literature synthesis covering twenty-five published works on machine-learning-based student performance prediction, and (b) an original quantitative experiment designed to empirically illustrate and test the central claims emerging from that literature — in particular, the relative predictive contribution of prior academic performance versus behavioural and demographic features.

3.2 Dataset

Because access to a live, restricted-distribution institutional dataset or to the original UCI Student Performance corpus of Cortez and Silva (2008) was not available within the constraints of this study, the empirical component uses a simulated illustrative dataset of 900 student records, generated to reproduce, in structure, the feature relationships consistently reported across the literature reviewed in Section 2. Continuous predictor variables — previous-term average score, attendance percentage, and weekly study time — were drawn from approximately normal distributions with means and standard deviations informed by typical ranges reported in prior empirical studies (e.g., Cortez & Silva, 2008). An ordinal parental-education variable, a count variable for past course failures, and two binary indicators (internet access, participation in extracurricular activities) were included to represent socio-demographic and behavioural features commonly used in the reviewed literature.

A continuous final-score outcome was generated as a weighted linear combination of these predictors, with previous-term performance assigned the dominant weight, plus Gaussian noise, reflecting the well-documented finding that prior academic performance dominates the explained variance in final outcomes (Cortez & Silva, 2008). A binary pass/fail label (pass defined as a final score of 50 or above) and a four-level grade band (F, C, B, A) were derived from the continuous

score for classification tasks. We emphasise that this dataset is illustrative rather than empirical fact about any real institution; it is used strictly to provide a transparent, fully reproducible computational demonstration of the comparative behaviour of different algorithms and feature sets under realistic, literature-informed assumptions, and all reported results in Section 4 were obtained by actually executing the analysis pipeline described below rather than being asserted post hoc.

3.3 Features

Seven predictor variables were used: previous-term average score (continuous, 0–100), attendance percentage (continuous, 0–100), weekly study time in hours (continuous), parental education level (ordinal, 1–4), number of past course failures (count), home internet access (binary), and participation in extracurricular activities (binary). These variables were selected because they correspond closely to the feature categories most frequently used across the twenty-five reviewed studies: academic history, attendance/behavioural engagement, and socio-demographic background.

3.4 Algorithms Evaluated

Consistent with the comparative emphasis of the reviewed literature, eight supervised learning algorithms were implemented using the scikit-learn library (Pedregosa et al., 2011): Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine (radial basis function kernel), k-Nearest Neighbours ($k = 9$), Gaussian Naive Bayes, and a Multilayer Perceptron artificial neural network with two hidden layers (16 and 8 units). For the continuous-outcome (regression) task, Ordinary Least Squares Linear Regression was used, consistent with its strong reported performance in prior work (e.g., the $R^2 = 0.8911$ result reported for linear regression on GPA prediction in a five-semester study of 853 students).

3.5 Data Partitioning and Preprocessing

The dataset was split into training (75%) and test (25%) subsets using stratified random sampling on the pass/fail label to preserve class balance, with a fixed random seed (42) to ensure reproducibility. Features used by distance-based and gradient-based algorithms (logistic regression, SVM, k-NN, and the neural network) were standardised to zero mean and unit variance using a scaler fitted on the training set only, to avoid information leakage from the test set. Tree-based algorithms (decision tree, random forest, gradient boosting) were trained on unscaled features, consistent with standard practice, since these algorithms are invariant to monotonic feature scaling.

3.6 Evaluation Metrics

For the classification (pass/fail) task, four standard metrics were computed on the held-out test set: accuracy (overall proportion of correct predictions), precision (proportion of predicted passes that were true passes), recall (proportion of actual passes correctly identified), and the F1-score (harmonic mean of precision and recall), which is particularly informative under class imbalance. For the regression (continuous score) task, root-mean-square error (RMSE) and the coefficient of determination (R^2) were used, consistent with the metrics reported in comparable prior studies.

3.7 Feature Importance and Ablation Analysis

To directly test the central literature finding that prior academic performance dominates predictive power relative to behavioural and demographic variables, two complementary analyses were performed. First, feature importance scores were extracted from a fitted Random Forest classifier, providing a ranked measure of each variable's contribution to classification accuracy. Second, an ablation study was conducted in which three nested feature sets — (i) academic history only, (ii) academic history plus behavioural variables, and (iii) the full feature set including demographic and other variables — were each used to train a Random Forest classifier under identical conditions, allowing direct comparison of the marginal predictive contribution of each feature category.

3.8 Literature Search and Selection Strategy

The narrative literature review in Section 2 draws on twenty-five peer-reviewed journal articles, indexed conference proceedings, and major systematic reviews identified through searches of IEEE Xplore, ScienceDirect, Springer Link, PubMed Central, and the arXiv preprint repository, using combinations of the search terms ‘student performance prediction’, ‘educational data mining’, ‘academic analytics’, ‘dropout prediction’, and ‘machine learning education’. Priority was given to (a) foundational and highly cited works that established the field's methodological baseline, (b) recent (2023–2026) studies reflecting the current state of the art, and (c) major systematic reviews and surveys capable of contextualising individual primary studies within the broader research landscape.

IV. RESULTS

4.1 Regression Results: Predicting Continuous Final Score

Ordinary least squares linear regression using the full feature set achieved a root-mean-square error (RMSE) of 6.012 points (on a 0–100 scale) and a coefficient of determination of $R^2 = 0.6784$ on the held-out test set. When the model was restricted to previous-term average score alone, performance declined only modestly, to $RMSE = 6.671$ and $R^2 = 0.6040$. This indicates that previous academic performance alone accounts for the substantial majority of explainable variance in final score, with the remaining six behavioural and demographic features jointly contributing an additional 0.074 to R^2 — a meaningful but clearly secondary improvement. This pattern is directly consistent with the findings of Cortez and Silva

(2008) and with the broader literature reviewed in Section 2, which consistently identifies prior academic performance as the dominant predictor of future outcomes.

4.2 Classification Results: Predicting Pass/Fail Outcome

Table 1 reports accuracy, precision, recall, and F1-score for all eight classifiers evaluated on the pass/fail prediction task. The Decision Tree classifier achieved the highest accuracy (0.9067) and F1-score (0.9466), narrowly outperforming Logistic Regression and Naive Bayes (both at 0.9022 accuracy, 0.9433 F1), and Random Forest (0.8978 accuracy, 0.9412 F1). The Multilayer Perceptron neural network achieved the lowest accuracy (0.8533) among the models tested, likely owing to the relatively small training sample (675 records) and the comparatively low-dimensional, low-noise feature space, which limits the advantage that high-capacity neural architectures typically demonstrate on larger or higher-dimensional datasets.

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	0.9067	0.9208	0.9738	0.9466
Logistic Regression	0.9022	0.9289	0.9581	0.9433
Naive Bayes	0.9022	0.9289	0.9581	0.9433
SVM (RBF kernel)	0.8978	0.9118	0.9738	0.9418
Random Forest	0.8978	0.9200	0.9634	0.9412
Gradient Boosting	0.8889	0.9235	0.9476	0.9354
K-Nearest Neighbors	0.8711	0.8750	0.9895	0.9287
Neural Network (MLP)	0.8533	0.9293	0.8953	0.9120

Table 1. Pass/fail classification performance across eight algorithms on the held-out test set ($n = 225$).

These results broadly corroborate the comparative findings reported elsewhere in the literature — for instance, the strong performance of random forest and logistic regression relative to neural networks under similarly structured tabular conditions, as similarly reported by the web-based dashboard study and the G-mean comparison study discussed in Section 2.3 — while also illustrating that the relative ranking of algorithms is sensitive to dataset size, dimensionality, and noise characteristics; no single algorithm dominates universally across all reported studies, a conclusion also reached by recent systematic reviews of the dropout-prediction literature.

4.3 Feature Importance Analysis

Table 2 presents feature importance scores extracted from a Random Forest classifier trained on the full feature set. Previous-term average academic score accounted for approximately 58 percent of total feature importance, more than three times the combined importance of attendance and study time, and roughly ten times the importance of any single demographic variable.

Feature	Random Forest Importance
Previous-term average score (prev_gpa)	0.5838
Weekly study time	0.1497
Attendance percentage	0.1389
Past course failures	0.0456
Parental education level	0.0434
Home internet access	0.0195
Extracurricular participation	0.0192

Table 2. Random Forest feature importance scores for the pass/fail classification task.

4.4 Ablation Analysis: Academic History versus Behavioural and Demographic Features

Table 3 reports the results of the nested ablation analysis. Using academic history alone (previous score and past failures) yielded an accuracy of 0.8978 and F1-score of 0.9415. Adding behavioural features (attendance, study time) did not improve, and very marginally reduced, both metrics (accuracy unchanged at 0.8978, F1 = 0.9412), and the further addition of demographic and other variables produced an identical result. This indicates that, within the present dataset and modelling

approach, almost all of the predictive signal recoverable by a Random Forest classifier is already captured by academic history variables alone, with behavioural and demographic variables contributing negligible additional discriminative power once a non-linear, interaction-capturing model is used.

Feature set (Random Forest)	# Features	Accuracy	F1-score
Academic history only (previous score, past failures)	2	0.8978	0.9415
Academic history + behavioural (adds attendance, study time)	4	0.8978	0.9412
Full feature set (adds parental education, internet, activities)	7	0.8978	0.9412

Table 3. Ablation comparison of nested feature sets using Random Forest classification.

It is worth noting that this finding does not imply that attendance, study habits, or socio-economic background are causally unimportant to learning outcomes; rather, it suggests that, statistically, much of their influence on final performance is already encoded within prior academic performance, which can be understood as a compressed summary of a student's accumulated behavioural and contextual circumstances up to that point in time — an interpretation consistent with the discussion offered by Cortez and Silva (2008) and reiterated in several subsequent studies.

V. DISCUSSION

5.1 Interpretation of Findings in Light of Prior Literature

The empirical results obtained in this study are highly consistent with the body of literature reviewed in Section 2. The dominance of previous academic performance as a predictor — accounting for roughly 58 percent of Random Forest feature importance and the majority of explained variance in the regression task — directly mirrors the original finding of Cortez and Silva (2008), in which first- and second-period grades alone accounted for over 50 percent of explained variance in final grades. It is also consistent with the five-semester GPA study reviewed in Section 2.2, in which linear regression using primarily historical score data achieved a very high R^2 of 0.8911, and with the consistent message across nearly all twenty-five reviewed studies that historical academic performance is the single most reliable feature category available to predictive models in this domain.

The comparative algorithm results obtained here — in which a relatively simple, interpretable Decision Tree model performed competitively with, and even marginally outperformed, more complex ensemble and neural methods — diverge somewhat from studies such as the random forest and G-mean comparison discussed in Section 2.3, which reported ensemble methods as clearly superior. This divergence is informative rather than contradictory: it illustrates a recurring theme in the broader EDM literature (Batool et al., 2023) that the relative ranking of algorithms is highly dependent on dataset size, feature dimensionality, noise level, and class balance, and that no single algorithm can be considered universally optimal for student performance prediction. Practitioners should therefore treat published rankings of algorithms as context-dependent evidence rather than as fixed prescriptions, and should empirically validate model choice on their own institutional data wherever possible.

5.2 Implications for Early Warning Systems

The finding that prior academic performance alone captures the large majority of available predictive signal carries an important practical implication: institutions seeking to build early-warning systems do not necessarily need to invest heavily in collecting and integrating granular behavioural or socio-demographic data before deploying a useful first-generation predictive tool. A model built solely on historical grades and attendance records — data that most institutions already routinely collect — can provide much of the achievable predictive value, which is an encouraging finding for resource-constrained institutions. This is consistent with the practical orientation of the enrolment-time dropout-prediction system described in Section 2.5, which demonstrated meaningful predictive value using only data available at the point of admission.

At the same time, the marginal but non-zero contribution of behavioural variables such as attendance and study time observed in the regression analysis (an additional 0.074 in R^2) suggests that richer, real-time behavioural data — such as that captured through learning management system logs, as discussed in Section 2.4 — retains value, particularly for refining predictions for individual students whose historical record alone does not clearly indicate risk status, or for shortening the lead time between behavioural change and detectable academic decline.

5.3 Methodological and Practical Limitations

Several limitations qualify the conclusions of this study. First, the empirical component relies on a simulated, literature-informed dataset rather than a real institutional or publicly archived dataset, because network access to retrieve such data was not available in the research environment used for this study; while the simulated data was deliberately constructed to reflect well-documented empirical relationships, results obtained from real student records may differ in magnitude, and ideally this analysis should be replicated on genuine institutional or benchmark data such as the original UCI Student Performance dataset (Cortez & Silva, 2008) as a direct follow-up. Second, the synthetic dataset assumes

approximately linear and additive relationships between predictors and outcome with Gaussian noise; real educational data may exhibit more complex non-linear interactions, sub-population heterogeneity, and non-Gaussian error structures, particularly across diverse demographic groups, which could alter the relative performance ranking of algorithms, especially favouring more flexible non-linear methods such as gradient boosting or neural networks on larger real-world samples. Third, the study addresses only structured, tabular features; it does not incorporate unstructured data sources (e.g., free-text assignments, classroom discourse, or multimodal sensor data) that are increasingly explored in the contemporary learning-analytics literature.

5.4 Ethical Considerations

The deployment of predictive academic models raises significant ethical considerations that merit explicit discussion, a point increasingly emphasised in the recent dropout-prediction literature reviewed in Section 2.5. First, predictive models trained on historical institutional data risk encoding and perpetuating existing structural inequities if certain demographic groups have historically been disadvantaged by factors unrelated to ability or effort; careful fairness auditing across subgroups is therefore essential before any operational deployment. Second, the use of such models to identify ‘at-risk’ students carries the risk of label-induced stigmatisation or lowered teacher and self-expectations if predictions are treated as deterministic rather than probabilistic and provisional. Third, given that predictive systems necessarily process sensitive personal and academic data, appropriate data governance, consent, and security frameworks must be established, consistent with applicable data protection regulation. The framing offered by the enrolment-time dropout-prediction study discussed in Section 2.5 — emphasising supportive outreach rather than punitive tracking — represents a useful ethical template for future deployments in this domain.

5.5 Directions for Future Research

Building on the findings and limitations identified above, several directions for future research appear promising. First, replication of the present comparative analysis on real, anonymised institutional datasets and on established public benchmarks would help validate the generalisability of the findings reported here. Second, the application of explainable artificial intelligence (XAI) techniques — such as SHAP value decomposition or counterfactual explanation — to performance-prediction models would help translate model outputs into actionable, interpretable guidance for educators, addressing a recurring critique in the literature that many high-performing models remain insufficiently interpretable for practical pedagogical use. Third, further work integrating structured academic-history data with unstructured behavioural data streams (e.g., LMS clickstream logs, as discussed in Section 2.4) using sequence-aware models such as recurrent neural networks or transformer-based architectures may improve early-prediction accuracy at earlier points in an academic term, when intervention value is highest. Finally, longitudinal studies evaluating whether interventions triggered by predictive models actually produce measurable improvements in student outcomes — moving beyond predictive accuracy toward causal, intervention-validated impact — represent an important and currently underexplored extension of this research area.

VI. CONCLUSION

This paper has examined the prediction of student academic performance from previous academic records through both a structured review of twenty-five published studies and an original, fully reproducible empirical evaluation. The literature review establishes that historical academic performance — prior grades, continuous assessment scores, and cumulative GPA — is consistently identified as the dominant predictor of future academic outcomes across diverse educational contexts, age groups, and modelling approaches, a conclusion that traces back to the foundational work of Cortez and Silva (2008) and that has been repeatedly corroborated in subsequent research spanning secondary and higher education, online learning environments, and dropout-prediction systems.

The empirical component of this study, conducted on a transparent, literature-informed simulated dataset, directly corroborates this conclusion: a linear regression model using previous academic performance alone achieved an R^2 of 0.604, compared with 0.678 for the full seven-feature model, while Random Forest feature importance attributed approximately 58 percent of total predictive contribution to previous academic performance alone. An ablation analysis further showed that the addition of behavioural and demographic features produced negligible improvement over an academic-history-only model under a Random Forest classifier, reinforcing the conclusion that prior performance functions, in practice, as a compact summary statistic capturing much of the information that behavioural and demographic variables would otherwise provide.

Among the eight classification algorithms evaluated, a simple, interpretable Decision Tree model achieved the highest accuracy (0.9067) and F1-score (0.9466), narrowly ahead of Logistic Regression, Naïve Bayes, Support Vector Machines, and Random Forest, while the neural network model evaluated performed comparatively weakly on this dataset — a reminder that algorithm selection should be guided by empirical validation on context-specific data rather than by reliance on universal claims of superiority drawn from other datasets.

Reference

1. Arifin, M., Widowati, W., Farikhin, F., & Gudnanto, G. (2023). A regression model and a combination of academic and non-academic features to predict student academic performance. *TEM Journal*, 12(2), 855–864.
2. Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. *arXiv preprint, arXiv:1606.06364*.
3. Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.

4. Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H.-Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28(1), 905–971.
5. Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.), *Proceedings of the 5th Annual Future Business Technology Conference (FUBUTEC 2008)* (pp. 5–12). EUROSIS.
6. Del Bonifro, F., Gabbrielli, M., Lisanti, G., & Zingaro, S. P. (2020). Student dropout prediction. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Proceedings, Part I* (pp. 129–140). Springer.
7. Discover Education (2025). Performance prediction using educational data mining techniques: A comparative study. *Discover Education*, 4, Article 89.
8. Dol, S. M., & Jawandhiya, P. M. (2023). Classification technique and its combination with clustering and association rule mining in educational data mining — A survey. *Engineering Applications of Artificial Intelligence*, 122, 106071.
9. Du, X., Yang, J., Hung, J.-L., & Shelton, B. (2020). Educational data mining: A systematic review of research and emerging trends. *Information Discovery and Delivery*, 48(4), 225–236.
10. Hasib, K. M., Rahman, F., Hasnat, R., & Alam, M. G. R. (2022). A machine learning and explainable AI approach for predicting secondary school student performance. In *Proceedings of the IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 399–405). IEEE.
11. IEEE (2024). Student academic performance prediction using machine learning with various features and scenarios. *IEEE Conference Publication*.
12. IEEE (2025). Student performance prediction system using machine learning. *IEEE Conference Publication*.
13. Knowles, J. E. (2015). Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. *Journal of Educational Data Mining*, 7(3), 18–67.
14. MDPI Information (2026). Machine learning and deep learning for dropout prediction in higher education: A review. *Information*, 15(3), Article 164.
15. Ozyurt, O., Ozyurt, H., & Mishra, D. (2023). Uncovering the educational data mining landscape and future perspective: A comprehensive analysis. *IEEE Access*, 11, 120192–120208.
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
17. Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.
18. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601–618.
19. Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs Data Mining and Knowledge Discovery*, 3(1), 12–27.
20. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355.
21. ScholarWorks UTRGV (2025). Predicting students' academic performance via machine learning models: An empirical review and practical application. University of Texas Rio Grande Valley, Faculty Publications and Presentations.
22. Wang, J., & Yu, Y. (2025). Machine learning approach to student performance prediction of online learning. *PLOS ONE*, 20(1), e0299018.
23. Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3), 547–570.
24. arXiv (2025). Evaluation of machine learning models in student academic performance prediction. *arXiv preprint*, arXiv:2506.08047.
25. arXiv (2025). Machine learning-driven student performance prediction for enhancing tiered instruction. *arXiv preprint*, arXiv:2502.03143.