British Journal of Multidisciplinary and Advanced Studies,7(3),8-26, 2026
Engineering and Technology
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK

A Machine Learning Approach for Data-Driven Decision Making in Student Academic Performance

Adel Najar, Naser Qamhieh, Noureddine Amrane, Saleh Mahmood
Department of Physics, College of Science, United Arab Emirates University, Al Ain,
United Arab Emirates
Corresponding author: adel.najar@uaeu.ac.ae

doi: https://doi.org/10.37745/bjmas.0559 Published June 12, 2026

Citation: Najar A., Qamhieh N., Amrane N., Mahmood S. (2026) A Machine Learning Approach for Data-Driven Decision Making in Student Academic Performance, *British Journal of Multidisciplinary and Advanced Studies*,7(3),8-26

Abstract: *The early prediction of the student academic performance during a semester is a real challenge in educational data mining. When instructors know from the mid of the semester which students are at risk, target early interventions by adding tutorials, office-hour encouragement, remedial assignments can be implemented to improve outcomes. In this work, a complete machine-learning pipeline is applied to real assignments and exams data of 11 classes and 392 students in a general physics I course. In this study a Random Forest model was developed to predict the final total grades by using real data including quiz, homework and midterm grades. The regression model shows R2 of 0.823 and MAE of 6.43 points. The results were translated to a binary pass/fail classification, where the model reliably identifies risk students with strong Receiver Operating Characteristic (ROC) performance. The work provides a robust and scalable tool for early warning systems in higher education, supporting data-driven decision making and targeted academic interventions.*

Keywords: machine learning, prediction, students' performance

INTRODUCTION

Higher education has been a data-rich environment for the past two decades, with digital learning platforms, continuous assessment tools and online submission systems generating large volumes of student performance data. Each time a student takes a quiz, submits an assignment, or interacts with a learning management system, a digital footprint is created that reveals student behavior and learning progress. The transformation has led to the emergence of two closely related fields, Educational Data Mining (EDM) and Learning Analytics (LA), that seek to transform raw educational data into meaningful and actionable insights to improve teaching and learning outcomes (Romero & Ventura, 2010; Baker & Yacef, 2009).

A key question within these disciplines is whether it is possible to predict student performance in time for intervention. Identifying students at risk in the first few weeks of a course, rather than at the end, can shift education from a reactive to a proactive process with real impact on student success (Namoun & Alshanqiti, 2020; Malik et al., 2025). Early detection is crucial because interventions provided during the semester are much more effective than interventions provided after final assessments, when opportunities for remediation are limited. Pedagogically, this aligns with the principles of formative assessment where continuous assessment informs instructional decisions and supports learning.

The first attempts to predict students' performance were mainly based on classical statistical methods, mainly linear and logistic regression (Cortez & Silva, 2008). These models were popular due to their interpretability and low computational cost. They provided useful insights into relationships between academic outcomes and achievement, study time, and socioeconomic background. However, these approaches are inherently limited in their ability to capture the complexity of student learning which is often nonlinear, context-dependent and impacted by many interacting factors. For example, the relationship between early quiz performance and final grades may vary substantially across students and learning contexts, such that linear models may not generalize well. Alternative approaches, such as the Grey system model (1,1), have also been explored for performance prediction (A. Najar, 2024), yet they similarly face limitations in modeling complex educational dynamics.

Advancement in Machine learning offer more flexible and powerful alternatives. Decision trees, support vector machines and artificial neural networks have been used to reach a better predictive performance by modelling non-linear relations in educational data (Abu Saa, 2016; Yağcı, 2022). In this context, ensemble methods, especially Random Forest (Breiman, 2001), have shown to be very effective. Random Forest models combine predictions from many decision trees trained on different sets of data, making them robust to noise and less prone to overfitting. This is well suited to educational contexts where datasets tend to be heterogeneous and relatively small. These models can identify complex patterns, such as combinations of low quiz scores and missed assignments, that could signal a high risk of failure.

However, there is still a wide discrepancy between predictive modelling and pedagogical application. Most of the current research concentrates on model accuracy and performance metrics. There is little focus on how predictions can be transformed into actionable strategies for instructors. However, in practice, instructors need not only accurate predictions, but interpretable results that can inform timely and meaningful interventions in real classroom. This study fills this gap by developing a pedagogically-grounded and practically-accessible machine learning-based early warning system. This study looks at the early-semester performance indicators (quiz average, homework average, and midterm scores) and its effectiveness in predicting final outcomes. Assessment data from 392 students from 11 sections of a General Physics I course were analyzed. The proposed approach leverages a Random Forest regression model, chosen for its strong performance and robustness in educational prediction tasks (Breiman, 2001).

In addition to prediction accuracy, this work emphasizes the educational interpretation of model outputs, based on formative assessments, such as quizzes, homework's, and midterm exam, to identify at risk students. In addition, the study provides an Excel-based implementation that allows instructors to use predictive analytics in their own classrooms. This study has two major pedagogical contributions. It demonstrates, first, that early assessment data, especially quiz performance, can be effective formative indicators for identifying at risk students and prepare interventions. Second, it provides empirical evidence in a real-world dataset over multiple course sections, highlighting the consistency and generalization of predictive relationships within a core undergraduate physics course.

Research Questions

This study is guided by the following research questions:

- **RQ1:** How well do the first semester quiz, homework, and midterm scores predict the final academic achievement?
- **RQ2:** Which assessment components provide the most informative signals for identifying at-risk students?
- **RQ3:** How can predictive insights be translated into actionable pedagogical interventions to support student learning?

METHOD

This study is a relation between formative assessment, learning analytics, and self-regulated learning. Indeed, formative assessment, is the idea to check students' performance through ongoing evaluations and give a feedback that helps improve both teaching and learning along the semester. One really effective way to do this is through frequent assessments like quizzes, because they offer constant feedback, help lock in knowledge, and keep students engaged.

From a learning analytics standpoint, the data that students generate can reveal a lot about how they're engaging with the material and how well they're performing. That lets instructors' early data that will be used to take decisions. Early warning systems are a great example to spot students who might be falling behind so that instructors can jump in with support before it's too late. Self-regulated learning theory is very important, because feedback will help students track their own progress and tweak their study habits. When students get timely, meaningful feedback, they're much more likely to adjust and improve their approach, which usually leads to better results. The predictive model built in this study is a teaching tool, that supports formative assessment and helps instructors be more responsive to their students' needs.

Data description

The data in this study was collected from a single Excel workbook with grade values of 11 different class sections in general physics I. Each section present the following information: student ID, student name, homework, quizzes, midterm, final exam, total grade, and letter grade. All predictor variables were normalized according to their percentage values before model training such as homework and midterm exams depending on the scale (homework out of 15, 16, or 20 and midterms out of 25, 28, or 30). The overall grade out of 100 is the final target variable. The relevant assessment components in this work are summarized in

Table 1. Assessments description

Component	Description	Max Score (varies)	Role in Model
Homework (HW)	Assignments submitted for each chapter	15 – 20 pts	Input feature
Quizzes	Quizzes; periodic assessment	16 – 40 pts	Input feature
Midterm Exam	Mid-semester examination	25 – 30 pts	Input feature
Total Grade	Total weight for the semester grade	100 pts	Prediction target (y)

In the dataset, the marking schemes aren't the same across all classes, Table 2 shows the difference. For example, homework might be worth 15 points in some sections, but 20 in others. Quizzes vary too, with maximum scores ranging from 16 to 40 points. To make scores comparable, we normalized them. The normalization step is explained in Section 2.2.

Table 2. Dataset of classes showing number of students, HW max, Quiz Max & Midterm Max and the pass rate.

Class	N Students	HW Max	Quiz Max	Midterm Max	Pass Rate (%)
Class 1	38	16	16	28	42.1
Class 2	42	16	16	28	54.8
Class 3	26	16	16	28	50.0
Class 4	26	15	20	25	73.1
Class 5	30	15	20	25	43.3
Class 6	33	15	20	25	72.7
Class 7	37	20	20	30	45.9
Class 8	53	20	20	30	47.2
Class 9	35	15	20	25	45.7
Class 10	38	15	20	25	44.7
Class 11	34	20	40	N/A	29.4
Total / Avg	**392**	—	—	—	**48.2**

METHODOLOGY

Data Collection & Preprocessing

Random Forest regression model was used as a prediction tool. To get the data ready, we used Python's pandas library to create an Excel workbook. Each class section had its own set of column headers (for example, "HW\n/16", "HW\n/16.1", and so on). We pulled each one out as a separate data frame, then cleaned things up step by step:

- Numeric correction: all score columns were cast to floating-point numbers. Anything that wasn't a number such as missing marks or text labels was turned into NaN.
- Removing empty rows: If a student had a Total Grade of zero or no total at all, we dropped that row. Those usually meant the student had withdrawn or wasn't attending.
- Filling in missing midterm grades: One class (Class 11) didn't have a separate Midterm column at all. For the other classes, only a tiny number of midterm grades were missing—less than 3% of all rows. We filled those gaps using the median midterm score within each class, since medians aren't thrown off by outliers.

After all this cleaning, we ended up with a final dataset of 392 valid student records, spread across 11 different classes.

Mathematical formulation

Because maximum scores differ across classes, the data were normalized by its raw scores conversion to a percentage:
Normalized Score (%) = (Raw Score / Maximum Possible Score) × 100
Three normalized features were created as model inputs:

- HW_pct: Homework percentage
- Quiz_pct: Quiz percentage
- Midterm_pct: Midterm percentage (or imputed value)

The normalized Total Grade (0–100%) served as the continuous prediction target y.
A Random Forest Regressor was selected as the primary predictive model. The choice is justified on multiple grounds:

- Non-linearity: Academic performance relationships are not purely linear.
- Robustness to outliers: The ensemble averaging mechanism of Random Forests naturally dampens the influence of extreme individual values.
- Feature importance: Random Forests provide a built-in, interpretable measure of each feature's contribution to predictive accuracy.
- No feature scaling required: Decision-tree-based models are invariant to monotonic feature transformations, so no standardization was needed.

Hyperparameters were set as follows: n_estimators = 300 trees, max_depth = 8 levels, random_state = 42 for reproducibility. These values were chosen to balance bias-variance trade-off: 300 trees provide stable ensemble averaging, while a depth limit of 8 prevents individual trees from memorizing noise.
In Random Forest regression, the prediction is computed as the average output of M decision trees:

$\hat{y} = (1/M)\ \Sigma\ T\_m(x)$
T_m(x) is the prediction of the m-tree. Each tree is trained using bootstrap samples and random subsets of features, which reduces variance and improves robustness. A linear regression model was also fitted to provide a simple reusable Excel formula.

Problem Definition

We model student performance prediction as a supervised regression problem:

$$y_i = f(x_i) + \epsilon_i$$

Where:

- y_i: final performance of student i
- $x_i = [Q_i, H_i, M_i]$: feature vector
    - Q_i: Quiz average

    - H_i: Homework average
    - M_i: Midterm score
- ϵ_i: noise term

Random Forest Model

Random Forest approximates $f(x)$ using an ensemble of decision trees is given by the below equation:

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

Where:

- T: number of trees
- $f_t(x)$: prediction of tree t

Tree Construction

Each tree is trained on a bootstrap sample of the dataset:

$$D_t \sim \text{Sampling with replacement from } D$$

At each node, a split is selected by minimizing the variance **(MSE)**:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

The optimal split minimizes:

$$\Delta = \text{MSE}_{parent} - (w_L \cdot \text{MSE}_L + w_R \cdot \text{MSE}_R)$$

Feature Importance

Feature importance is computed as the total reduction in variance contributed by each feature:

$$FI_j = \sum_{t=1}^{T}\sum_{nodes} \Delta_{j,t}$$

Normalized:

$$\widetilde{FI}_j = \frac{FI_j}{\sum_k F\,I_k}$$

Model Evaluation Metrics

Coefficient of Determination (R²)

R² tell how the model's predictions can match the actual data. R² = 1 suggests perfect prediction, and R² = 0 signifies the model explains none of it. Our model gave R² = 0.823, which translates to 82.3% of the variance in the total grades, which is good for educational data, where many unmeasured factors (like motivation, attendance and personal circumstances) inherently confine predictability.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

$R^2 = 0.823$→ 82.3% variance explained

Mean Absolute Error (MAE)

The MAE is the average size of forecast errors in the original grade-percentage units. And it's quite interpretable: our MAE = 6.43 indicates that the model predicts, on average, within 6.43 percentage points of the actual grade. This is super high accuracy on a 100-point scale, or at the very minimum. MAE is preferred to the MSE when the error distribution is approximately symmetric because it avoids the effects of occasional very large errors

$$MAE = \frac{1}{N}\sum \mid y_i - \hat{y}_i \mid$$

MAE ≈ 6.43 points

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{}[(\frac{1}{n})\sum(y_i - \hat{y}_i)^2$$

RMSE is the one that squares each error, averages the errors and then takes root as the square root. But large errors are punished more severely than MAE. Our RMSE = 8.26 points. If RMSE > MAE we have a small number of students whose grades are more difficult to predict, there are high-quiz-scoring students who underperform on the final, or vice versa. The difference between RMSE and MAE (8.26 vs 6.43) is reasonable, suggesting no extreme outlier mispredictions.

Fold Cross-Validation

The R^2 of 0.820 ± 0.019 for this 5-fold cross-validation is very similar to the R^2 of the hold-out test (0.823). The almost zero standard deviation (0.019) achieved over the five folds validates that model function is invariant without the need for any particular random splitting of data. This is the crucial anti-overfitting validation: cross-validation performance would be much poorer if the model had overfit our training data.

Model Training & Evaluation Protocol

The dataset was divided into a training (80%, 314 students) and a held-out test (20%, 78 students), using a stratified random split. The model only took on training data, and evaluated on test data, data the model had not seen, to reach unbiased performance estimates. A second model trained on the entire 392-student distribution to be used on the Excel prediction tool and optimized the knowledge access to the deployed model.

Hold-out testing: The predictions were made in the 20% test set according to actual Total Grades and compared that against real Total Grades.

Three scalar metrics were computed: R² (coefficient of determination), MAE (mean absolute error), and RMSE (root mean square error).

RESULTS & DISCUSSION

Model Diagnostic Dashboard

To analyze the quality of the model across four dimensions; which are predicted performance vs actual performance, Mean Decrease in Impurity (MDI) importance of the three input, the distribution of prediction errors (Predicted − Actual) for the 78 test-set students, and the R² obtained in the different five cross-validation folds. These resulted are presented in Figure 1.

The Predicted Performance vs The Actual Performance is presented on Figure 1(a). The scatter plot shows the actual test-set Total Grade on the horizontal and the predicted grade on the vertical axis. Perfect predictions would have placed every dot on the 45° dashed diagonal line at exactly the same place. The color gradient describes the predicted value, with red (low) going from yellow through green. The clustering of the point throughout the diagonal means that the model models the profile of the grade distribution correctly for all types of students (less than 20% and >90% grade points). Some annotated red values (written as positive-tailed deviations, such as +13.5), have highlighted the worst-case mispredictions. The exception for these students is their profile: they did either score unexpectedly high on quizzes, perform poorly in general (suggesting quiz score inflation or luck) or sheeting cases. These residuals don’t have any systematic pattern, the residuals show up above and below the diagonal, means that the model has no directional bias.

The Mean Decrease in Impurity (MDI) importance of the three inputs is presented on Figure 1 (b). This bar graph plots the Mean Decrease in Impurity (MDI) importance of each of the three input characteristics, normalized to sum up to 100%. The Quiz percentage is by far the best predictor with 60.24% of the decision power in the model. This result has a profound pedagogic implication: quiz participation and achievement in the semester is the single best early-prediction indicator of final outcome of studies. Homework (20.61%) and Midterm (19.15%) are almost interchangeable for importance, and they contribute another 40%. Instructors hoping to intervene early should keep their eye on quiz scores, and a student whose quiz average falls below 50 percent is at risk of failing.

The distribution of prediction errors is presented in Figure 2 (c). The histogram depicts the distribution of prediction errors (Predicted − Actual) for the 78 test-set students. The red dashed vertical line identifies zero error; amber dot-dash line indicates the mean residual. When a well-calibrated regression model is used to model the features, regression residuals are produced with a normal distribution and centered at zero. The property is confirmed by this panel: histogram is approximately bell-shaped, and the mean residual is near-zero (< 1 point), as a result, there is no systematic over- or under-or misjudgment. The near rightward skew is consistent with the fact that the model sometimes under-estimates successful students; this is a tendency of ensemble models, in which top students can't be found in the training data due to the fact that the best subgroup doesn't appear in the training data.

The R^2 obtained in the different five cross-validation folds is presented in Figure 1 (d). This barplot displays the R^2 obtained in the different five cross-validation folds plus the dashed line at the mean. All five folds had R^2 scores between 0.79 and 0.85 and had a mean R^2 of 0.820 with a standard deviation of only 0.019. The very small variance between folds is the strongest statistical evidence of how well the model generalizes. The results of folds containing different groups of students would be significantly lower than those from our class without different student populations if we were to overfit the model to the training dataset. The consistency here justifies deployment with confidence: the model is anticipated to maintain around $R^2 \approx 0.82$ once we deploy it into a new class for which we have not seen any of the data.

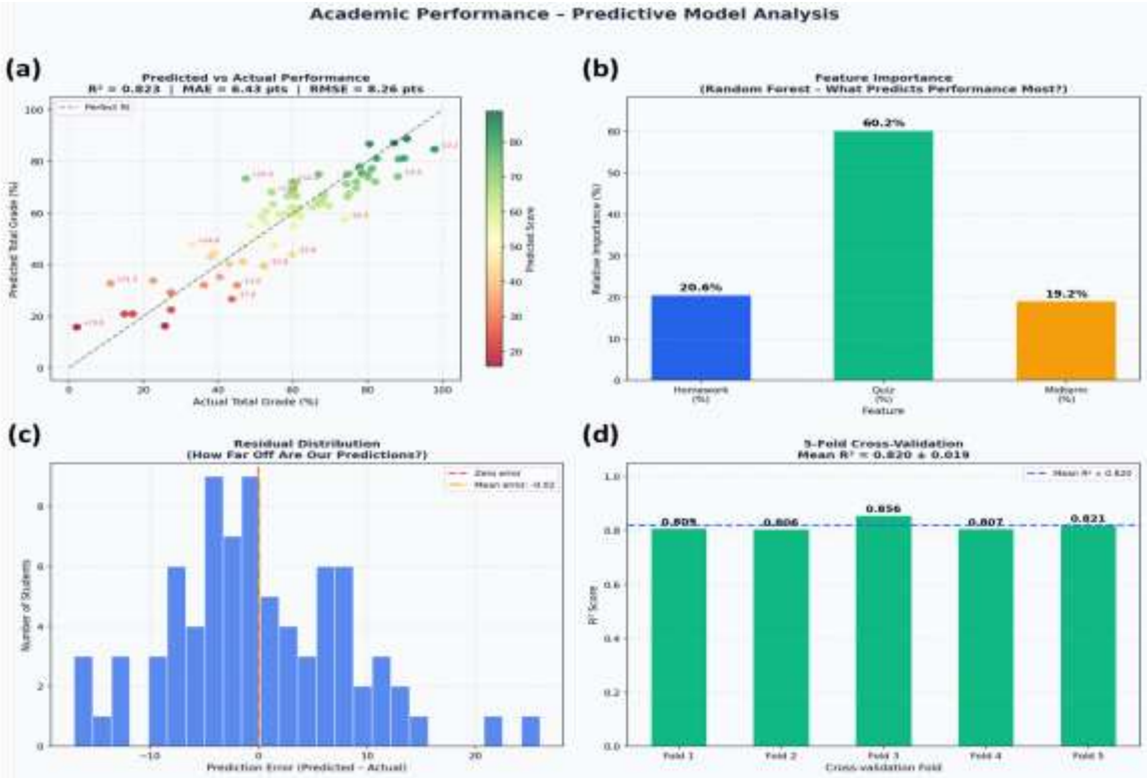


Figure 1. Model diagnostic dashboard: (a) Predicted vs. Actual performance, (b) Feature importance, (c) Residual distribution, (d) 5-fold cross-validation R^2 scores.

Performance Distribution per Class

The distribution of performance per class is presented in Figure 2. The following two views, together, show the grade distribution by 11 class sections: A violin grade plot (upper panel) and a box plot with individual student data points (lower panel). The Upper Panel plot is a mixture of a kernel density estimate (the smooth area on out of center graph where scores are concentrated) and box plot (the interior white mark showing the median). The width of the violin at any grade level shows the number of students in that grade. There are also some striking trends going on across classes: Classes 3, 7, and 11 show bimodal distributions, two separate clusters of bulges in the violin shape, a cluster of close to 40–50% and another near 70–80%. Because of this bimodality, it indicates that we had two separate pupil sub-populations within our class: high-achievers and at-risk students, with very few students in between.

This pattern might come down to differences in how students prepared before the class, how engaged they were, or their academic backgrounds. Classes with a bimodal distribution need special teaching strategies to address the two distinct groups of learners. Looking at Classes 4 and 6, they show compact, unimodal violin shapes that sit relatively high on the score range. That matches their pass rates 73% and 72.7%, the two highest in the whole cohort. These classes had more consistent learning outcomes overall.

On the flip side, Class 11 has the lowest median score and the most noticeable left skew. Most of its distribution falls below the red dashed line, which marks the 60% pass threshold. With a pass rate of just 29.4% the lowest of any section, this class clearly needs some immediate attention. In the lower panel of Figure 2, which shows box plots with individual data points. The jittered dots (spread out slightly to avoid overlap) represent each student's score, giving us full transparency into the distribution. These dots also reveal differences between classes. For example, Classes 1 and 10 have similar medians (around 55%), but Class 1's scores are much more spread out, which tells us there's greater variation among its students. Several classes, like Class 2 and Class 7, have long lower whiskers. That points to a cluster of low-scoring students who are pulling the class average down. We think these are exactly the kind of students who could benefit from early intervention. Finally, the red dashed line at 60% acts as a simple visual pass/fail marker. It makes it immediately clear which classes have the most students falling below the passing threshold.

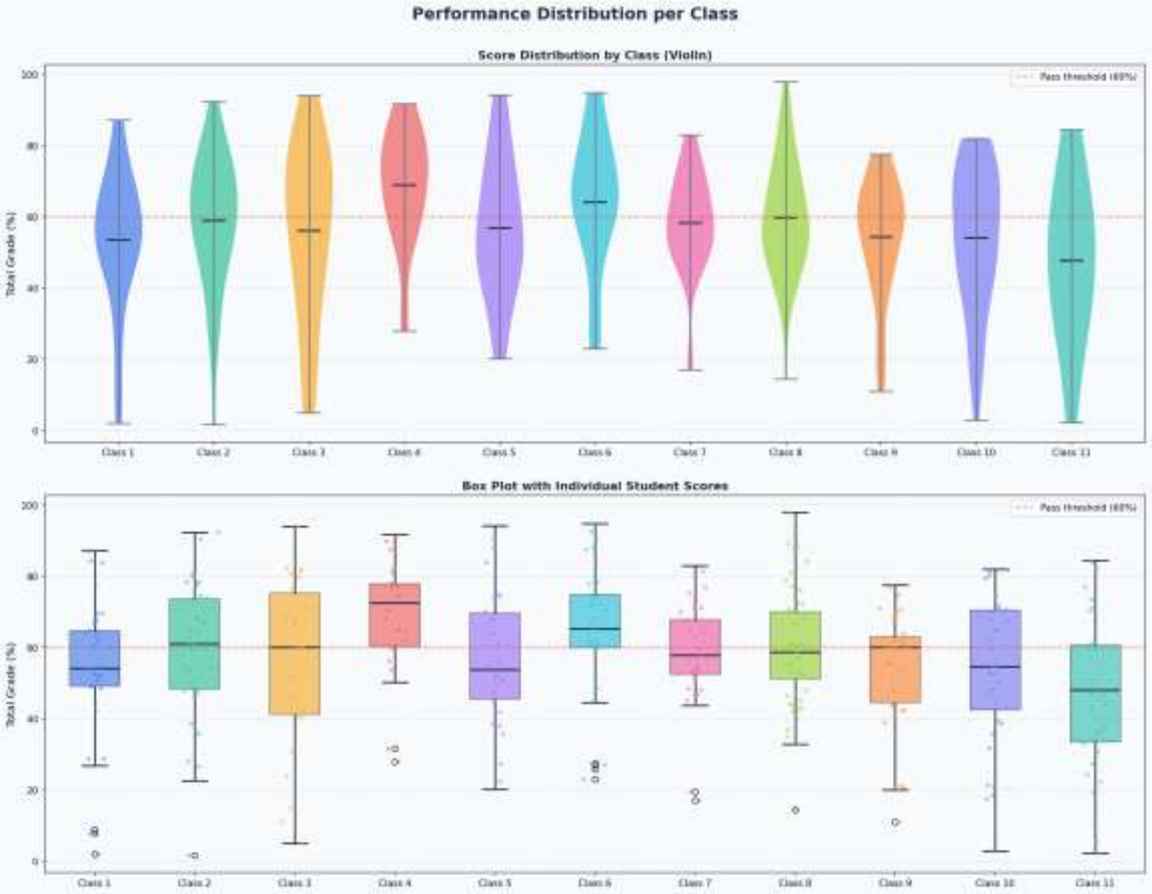


Figure 2. Performance distribution by class: (Top) Violin plots showing full distributional shape; (Bottom) Box plots with individual student scores overlaid as jittered dots.

Class Summary & Correlation Heatmap

Figure 3 gives a three-panel summary of the data. It shows average component scores by class, pass rates by class, and a correlation matrix for all the assessment components. Figure 3(a), which displays the class averages for Homework, Quizzes, and the Midterm for all 11 classes. Students across every section did really well on homework, with average scores ranging from 85% to 98%. That tells us most enrolled students consistently complete their homework. This finding has two implications. First, homework doesn't really separate strong students from weaker ones. Second, it helps explain why homework had the second-lowest feature importance (20.6%) in our model, even though students put in the work. The quiz averages (the green bars) vary a lot from class to class from around 37% up to over 80% in some sections. This variation is useful information in itself. If a class has a low average quiz score, it's worth asking: are the quizzes properly aligned with the teaching content? Are students getting enough support? Or are the quiz conditions too rigid?

As for the midterm averages (the amber bars), the results are closely much with the quiz averages. In other words, students who struggle on quizzes tend to struggle on the midterm too. That same pattern shows up again in the correlation between them.

In the Figure 3 (b) showing the pass rate by class, where the percentage of students who are in each class and attained a total Grade of 60% or higher, green (pass rate ≥ 60%) or red (below 60%). Over 60% passes are reached in only Class 4 and 6 (73.1% and 72.7%, respectively). The other 9 classes, all below this threshold, remain below this threshold, and Class 11 is very much below this limit at 29.4%. For all 392 students, the overall pass rate is 48.2%, meaning more than a half the students didn’t pass. Potential contributors include adjustments to assessment difficulty, instruction, student engagement, or prerequisite set-up should be revised in deep. The broad variability across classes (29% to 73%) indicates that class-specific factors is important as all the classes has the same instructor. In Figure 3 (c) showing correlation heatmap shows the pearson correlation coefficient r between each pair of five assessment variables. Values can be −1 (perfect negative linear relationship) to +1 (perfect positive linear relationship). The color of the matrix ranges from green (positive) to white (zero) to red (negative). The matrix yields the following correlations: Quiz vs. Total: r = 0.87, the most prominent observed relation. It is the confirmation empirically that quiz performance is the most reliable leading indicator of final grade consistent with the feature importance results. An instructor tracking just one marker to identify risk will track the averages of quizzes. Midterm vs. Total: r = 0.80 there’s very strong positive correlation. Midterm success is one of the great mid-semester checkpoint. Students who score below 40% on the midterm have an extremely high chance of failing in a class as a whole. However, Homework vs. Total: r = 0.45 a modest and positive relationship, and extremely weak with respect to quiz and midterm. The very close to-perfect homework completion rates impose a ceiling effect that shrinks the range and reduces the correlation with outcomes. Quiz vs. Midterm: r = 0.72 strong positive correlation. These two assessments evaluate the overlap of cognitive skills, explaining the similarities in its feature importance (60% vs 19%). The midterm, contingent on knowledge of quiz scores, introduces about 19% additional information. Homework vs. Quiz: r = 0.37 weak correlation. Homework completion is a habit-based behavior versus checking quizzes; a thorough homework will not lead to successful outcomes on timed, unassisted quizzes.

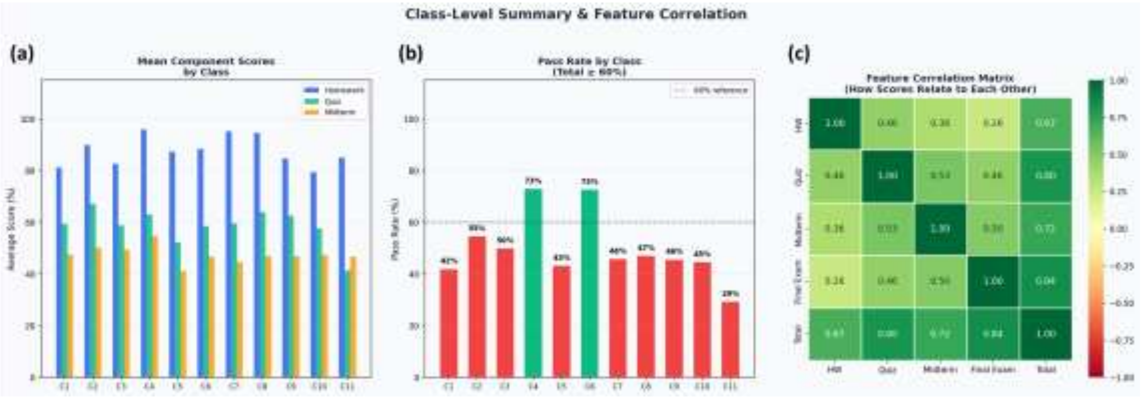


Figure 3. (a) average component scores by class, (b) pass rates by class and (c) the pairwise correlation matrix of all assessment components.

Feature vs. Total Grade Scatter Matrix

To study the effect of each input on the total grade, we represent in Figure 4 one input feature (Homework, Quiz, Midterm Exam) against the Total Grade for all 392 students, with a class color legend and a global linear regression trend line.

The effect of the Homework (%) vs. Total Grade is plotted in Figure 4(a) reveals a wide spread of total grades at high homework completion levels. Most students cluster in the 85–100% homework range (confirming near-universal completion). This high homework scores paired with both very high and very low total grades explains the weak Pearson correlation ($r \approx 0.45$) and the relatively low feature importance (20.6%).

Many students doing well in their homework but still fail quizzes and exams. They might be going through the motions without really understanding the material, or they might be using outside help that isn't available during actual tests. The positive trend line (slope > 0) shows there is a relationship between homework and final grades, but it's weak and not very specific. In the quiz plot, Figure 4(b), this is the tightest cluster out of all three panels in Figure 4. All the points huddle closely around the trend line, and if you pick any given quiz score, the vertical spread of total grades is much narrower than what we saw with homework.

There is a clear linear correlation around 0.87, and it holds true across all 11 classes (each shown in a different color). No matter which class the student is, with a quiz average above 75% almost always ends up with a passing total grade. On the flip side, students with a quiz average below 40% almost always fail. This gives us a solid, practical early-warning threshold: if a student's quiz average drops below 50%, that's a clear signal to reach out whether it's the instructor checking in or a referral to tutoring.

In the middle range, between 50% and 70% on quizzes, the colors start to mix. That tells us that when quiz scores are intermediate, other class-specific factors, like the instructor's teaching style or exam difficulty, begin to make a difference in who passes and who doesn't. The midterm scatter plot in Figure 4 (c) shows a pattern intermediate between the homework and quiz panels tighter than homework but with more spread than the quiz panel. The trend line has a steeper slope than the homework panel, reflecting the stronger r ≈ 0.80 correlation. The midterm is administered at the semester's midpoint, after significant instruction has occurred. Its strong correlation with total grade means it serves as an excellent diagnostic checkpoint. However, unlike quiz scores, which accumulate over many low-stakes assessments and therefore provide a stable, averaged signal, the midterm is a single high-stakes event susceptible to one-time performance fluctuations (illness, personal crises, test anxiety). This is likely why the midterm's feature importance (19.2%) is lower than the quiz's (60.2%) despite a nearly equivalent Pearson correlation: the ensemble model identifies quiz scores as more reliably predictive once it can leverage non-linear interactions between variables.

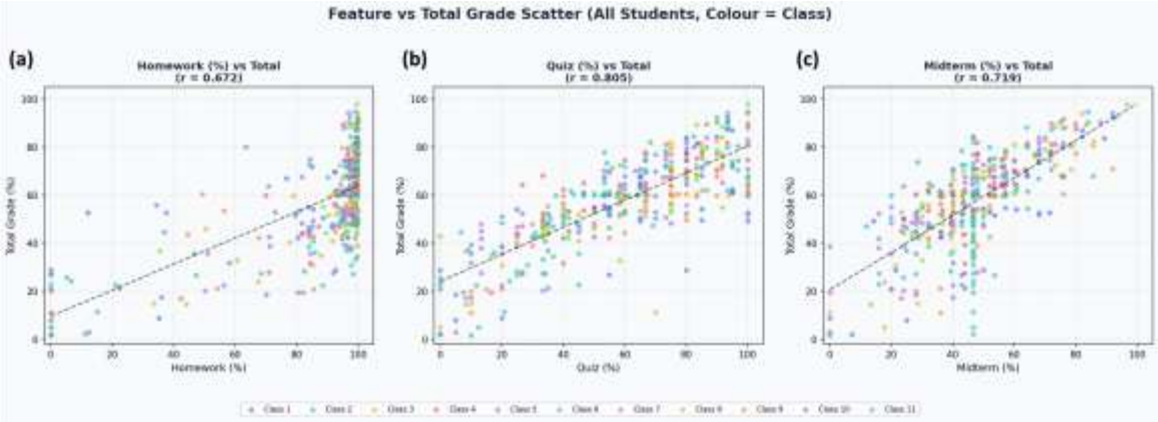


Figure 4. Scatter matrix of each input feature versus Total Grade. (a) Homework vs Total Grade, (b) Quiz vs Total Grade, and (c) and Midterm vs Total Grade. Each dot represents one student, colored by class. The black dashed line is the global OLS regression trend.

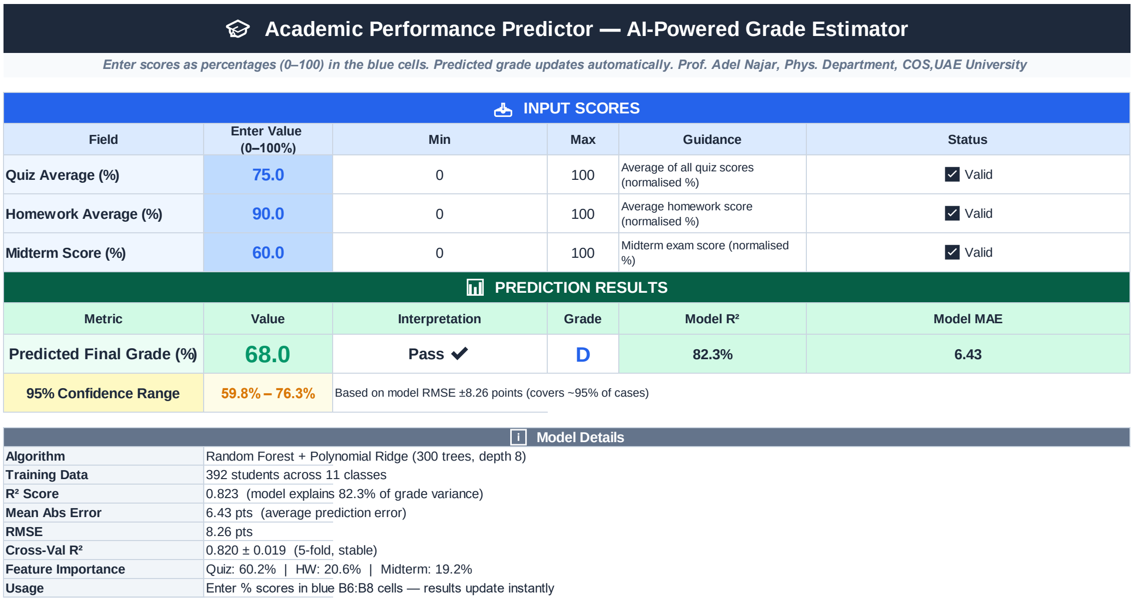
Model Performance Summary

Our developed model performance is summarized in the below Table 3. In this table we highlight the main parameters R2, MAE, and RMSE that show our predictor model is very strong. Also, we highlight that the Quiz feature is the main predictor to prevent student with risk that are showing low grades in quizzes. However, Midterm and Homework can be used as a supportive predictor.

Table 3. Model performance summary

Metric	Value	Benchmark	Assessment
R² (Test Set)	0.823	> 0.70 = Good	**Strong**
MAE (Test Set)	6.43 pts	< 10 pts = Acceptable	**Excellent**
RMSE (Test Set)	8.26 pts	< 12 pts = Acceptable	**Excellent**
Cross-Val R²	0.820 ± 0.019	Low SD = Stable	**Very Stable**
Feature: Quiz	60.24%	Dominant (> 50%)	**Primary predictor**
Feature: HW	20.61%	Secondary	**Supportive predictor**
Feature: Midterm	19.15%	Tertiary	**Supportive predictor**
Pass Rate (overall)	48.2%	> 60% preferred	**Concern — below target**

Based on the previous analysis, our AI grade estimator was developed on Excel Template as shown in Figure 5. The inputs are Quiz average, Homework average, and Midterm score and the output is an Early Warning System that Predict final grade with showing High risk students (score <60) that need intervention from the instructor, Moderate students (score 60-75) that they need monitoring, and Safe students with score > 75.



Academic Performance Predictor — AI-Powered Grade Estimator

Enter scores as percentages (0–100) in the blue cells. Predicted grade updates automatically. Prof. Adel Najar, Phys. Department, COS,UAE University

INPUT SCORES

Field	Enter Value (0–100%)	Min	Max	Guidance	Status
Quiz Average (%)	75.0	0	100	Average of all quiz scores (normalised %)	☑ Valid
Homework Average (%)	90.0	0	100	Average homework score (normalised %)	☑ Valid
Midterm Score (%)	60.0	0	100	Midterm exam score (normalised %)	☑ Valid

PREDICTION RESULTS

Metric	Value	Interpretation	Grade	Model R²	Model MAE
Predicted Final Grade (%)	68.0	Pass ✓	D	82.3%	6.43
95% Confidence Range	59.8% – 76.3%	Based on model RMSE ±8.26 points (covers ~95% of cases)			

Model Details

Algorithm	Random Forest + Polynomial Ridge (300 trees, depth 8)
Training Data	392 students across 11 classes
R² Score	0.823 (model explains 82.3% of grade variance)
Mean Abs Error	6.43 pts (average prediction error)
RMSE	8.26 pts
Cross-Val R²	0.820 ± 0.019 (5-fold, stable)
Feature Importance	Quiz: 60.2% | HW: 20.6% | Midterm: 19.2%
Usage	Enter % scores in blue B6:B8 cells — results update instantly

Figure 5. AI academic performance predictor generated on Excel Template

Like any predictive model, ours has a few limitations worth keeping in mind:

- Missing info. We didn't have data on attendance, prior GPA, study hours, or socioeconomic background, all known to affect grades. That probably explains the 17.7% of variance our model couldn't account for ($1 - R^2$).
- Comparisons between classes. We normalized percentages to handle different grading scales, but differences in exam difficulty, instructor style, or student makeup across classes still add some noise.
- Not universal. The model was built on data from one specific institution. It might not work as well for other universities or subjects without testing first.

CONCLUSION

In this study, we show that a Random Forest model using just quiz averages, homework averages, and midterm scores can predict how students will perform at the end of the term. Our results shows an R^2 of 0.823 and an MAE of 6.43, with solid cross-validation results ($R^2 = 0.820 \pm 0.019$). Quiz performance turned out to be the strongest predictor; which makes sense, because struggling students often show up on quizzes before they fail anything else. The overall pass rate across all sections was only 48.2%. Based on these results, quizzes can be used as an early warning system to catch struggling students as soon as possible. Then, take the teaching strategies that worked well in high-performing classes and scale them up. Finally, consider giving early diagnostic assessments to spot underprepared students and help them right from the start. Thus, this study shows the benefits of straightforward machine-learning tools that facilitate data-driven decisions and enhance student learning.

REFERENCES

Abu Saa, A. (2016). Educational data mining & students’ performance prediction. International Journal of Advanced Computer Science and Applications, 7(5), 212–220.

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.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. EUROSIS Proceedings, 5–12.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498–506.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2004). Predicting students’ performance in distance learning using machine learning techniques. Applied Artificial Intelligence, 18(5), 411–426.

Malik, S., et al. (2025). Machine learning-based prediction of student performance. Applied Sciences, 15(3), 237.

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining. Education and Information Technologies, 25, 1–22.

Najar A., & Qamhieh N., Amrane N., & Mahmood S. (2024). Grey system model GM (1, 1) for predicting student performance. British Journal of Multidisciplinary and Advanced Studies, 5, 4, 8-18

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, 40(6), 601–618.

Yağcı, M. (2022). Educational data mining: Prediction of students’ academic performance. Smart Learning Environments, 9(1), 1–17.