The Indonesian Journal of Computer Science
www.ijcs.net
Volume 15, Issue 2, April 2026
https://doi.org/10.33022/ijcs.v15i2.5125

An Integrated Machine Learning Framework for Early Prediction of Student Academic Performance in University Management Systems

Omeed Mustafa Mohammed
omeed.mohammed@nawroz.edu.krd
Nawroz University

Article Information

Received: 10 Apr 2026
Revised: 22 Apr 2026
Accepted: 27 Apr 2026

Keywords

student prediction; machine learning; EDM; analytics; UMS

Abstract

Forecasting student academic standing at an early stage is a critical challeng in higher education. This paper presents a supervised machine learnin classification framework applied to final examination records from 1,01 students across the English and Translation Departments of Nawro University, Duhok, Kurdistan Region, Iraq, spanning Semesters 1–6 under th Bologna Process. Six classifiers were evaluated using stratified five-fold cross validation and an independent held-out test set. Each student's six course fina marks served as input features; the output was one of four GPA-derived tier (Excellent, Good, Satisfactory, Poor/At-Risk). SVM (RBF) achieved th strongest performance: 96.06% test accuracy, 94.09% cross-validate accuracy, and 92.55% macro F1-score. Results show that a lightweight M pipeline using only routine assessment data can exceed 94% predictio accuracy, making it a viable early-warning component within existin University Management System infrastructure.

A. Introduction

Institutions of higher education are generating unprecedented volumes of academic data through their management and information systems, yet the potential of this data to proactively support student success remains substantially underexploited. Across Bologna Process universities — which govern curriculum structure and credit assignment according to standardised European frameworks, a model now adopted broadly in Iraqi higher education — semester examination records are collected with regularity and precision but are rarely fed back into decision-support tools that could alert academic staff to struggling students before failure occurs.

In recent years, the fields of Educational Data Mining (EDM) and Learning Analytics (LA) have developed into important research domains within computer science and data analytics, providing a robust body of theory and technique for transforming institutional records into actionable predictive models [1, 2]. However, a persistent observation in this literature is that most published prediction systems exist only as research artefacts and never transition into operational deployment within the actual administrative systems universities use [3]. The consequence is a structural disconnect: research advances accumulate rapidly while the students who could benefit from early identification and outreach remain underserved.

This work addresses that disconnect directly. We present an end-to-end ML classification framework grounded in real examination data from the English and Translation Departments of Nawroz University's Language College, spanning Semesters 1 through 6 under the Bologna credit system. By restricting features to the six course-level final marks that department administrators collect as a matter of routine, the framework imposes zero additional data-collection burden — a deliberate design choice intended to maximise the likelihood of real-world adoption within existing University Management System (UMS) infrastructure.

The specific contributions of this work are: (1) A fully operational ML prediction framework applied to two departments of Nawroz University's Language College, classifying 1,015 students into four academic standing tiers from six course marks. (2) A rigorous six-algorithm benchmarking study using both stratified cross-validation and independent held-out testing, yielding a peak test accuracy of 96.06% with SVM (RBF). (3) Course-level feature importance analysis revealing a balanced importance distribution (C5=0.193, C4=0.178, C1=0.172, C3=0.161, C2=0.158, C6=0.138). (4) A contribution to the underrepresented EDM literature covering Iraqi Kurdistan's higher education sector.

Based on the identified research gaps, this study investigates whether student academic performance tiers can be accurately predicted using only routinely collected course-level examination marks within a university management system environment.

B. Research Method

Framework Architecture

The proposed framework comprises five sequential operational stages: data ingestion from official examination registries; record validation and cleaning; derivation of prediction-ready feature vectors and GPA performance tier labels; multi-algorithm training, cross-validation, and held-out testing; and translation of model outputs into actionable risk signals suitable for UMS dashboards. A deliberate design constraint is that no data source beyond

final examination marks is required, ensuring the framework can be activated immediately upon completion of each semester's assessment process.

Study Dataset

The dataset for this study was obtained from the English and Translation Departments within the College of Languages at Nawroz University in Duhok, Kurdistan Region, Iraq. It covers all six semesters of the undergraduate program, following the Bologna Process and ECTS framework. Each student record includes final examination marks for six courses.

Initially, the dataset consisted of 1,037 records. However, 22 records were excluded because they contained all-zero marks, which likely indicate student non-attendance. After this filtering step, a total of 1,015 valid records remained for analysis. Of these, 560 records were collected from the English Department, while 455 records were from the Translation Department. A detailed overview of the dataset is presented in Table 1.

Table 1. Dataset Characteristics — Language College, Nawroz University (n=1,015)

Attribute	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	Sem 6	Combined
Raw records	179	180	131	169	200	178	1,037
Excluded (absent)	0	11	0	6	0	2	19
Usable records	176	169	131	163	200	176	1,015
Courses per semester	6	6	6	6	6	6	6
Train/test partition	-	-	-	-	-	-	80% / 20%

Data Cleaning and Validation

Five preprocessing steps were carried out to prepare the dataset for analysis. First, records in which all six course marks were zero were identified as representing absent students and were therefore removed (19 records). Student identifiers were then standardized to eliminate potential duplication. Any non-numeric entries were converted to zero to ensure consistency across the dataset.

In addition, three records containing missing values for individual course marks were retained, with the missing entries replaced by zero. Finally, all calculated GPA values were verified to ensure they fell within the Bologna-compliant range of 0.00 to 4.00.

Performance Tier Labels

The classification target was defined as a four-level academic performance tier based on each student's semester GPA, in accordance with Bologna Process conventions. Across the full dataset of 1,015 records, the distribution of students was as follows: Satisfactory (n = 458, 45.1%), Good (n = 335, 33.0%), Poor/At-Risk (n = 182, 17.9%), and Excellent (n = 40, 3.9%). The detailed tier classification schema is presented in Table 2.

Table 2. GPA Performance Tier Schema (Bologna Process, Nawroz University)

GPA Range	Performance Tier	Combined (n=1,015)

3.50 – 4.00	Excellent	40 (3.9%)
2.50 – 3.49	Good	335 (33.0%)
1.50 – 2.49	Satisfactory	458 (45.1%)
0.00 – 1.49	Poor / At-Risk	182 (17.9%)

Feature Construction

Each student's feature vector was composed of six individual course final marks (C1–C6), treated as continuous numerical variables on a 0–100 scale. Composite features, such as the total marks (sum) and arithmetic average, were deliberately excluded to prevent target leakage.

The dataset of 1,015 records was then divided into two subsets: a training set comprising 80% of the data (n = 812) and an independent held-out test set comprising the remaining 20% (n = 203). Stratified sampling was applied to ensure that the distribution of performance tiers was preserved across both subsets.

Classifiers Under Evaluation

Six classifiers were included: Decision Tree (max depth 5), Random Forest (100 trees), Gradient Boosting (100 estimators), Support Vector Machine with RBF kernel (C=1.0), K-Nearest Neighbors (k=5, Euclidean distance), and Naive Bayes (Gaussian). All models were implemented in Python 3.11 using scikit-learn.

Evaluation Design

Model performance was evaluated using two complementary strategies: stratified five-fold cross-validation and an independent held-out test set (n = 203), which was not used during model training. Under both evaluation protocols, four performance metrics were calculated: overall classification accuracy, along with macro-averaged precision, recall, and F1-score.

Macro averaging was chosen to ensure that the model's performance on minority classes—Excellent (n = 17) and Poor/At-Risk (n = 31)—was given equal importance alongside the majority class, Good (n = 70).

C. Result and Discussion

Figure 1 illustrates the distribution of students across the four GPA performance tiers, disaggregated by department and semester. The Satisfactory tier represents the largest proportion of students (n = 458, 45.1%), followed by Good (n = 335, 33.0%), Poor/At-Risk (n = 182, 17.9%), and Excellent (n = 40, 3.9%).

The relatively small proportion of students in the Excellent tier highlights the importance of using macro-averaged evaluation metrics, as these ensure that each performance tier is given equal weight regardless of its sample size.

Table 3 and Figure 2 present the benchmarking results across all six classifiers. Among the evaluated models, the SVM with an RBF kernel achieved the highest cross-validated accuracy (94.09%), along with the strongest macro F1-score (92.55%) and macro precision (94.85%), establishing it as the best-performing model overall.

On the held-out test set, SVM also achieved the highest accuracy (96.06%), followed by Random Forest (92.12%) and KNN (90.15%). In contrast, the Decision Tree model yielded the lowest performance across all evaluation metrics, suggesting that single-tree approaches lack sufficient representational capacity for this classification task.

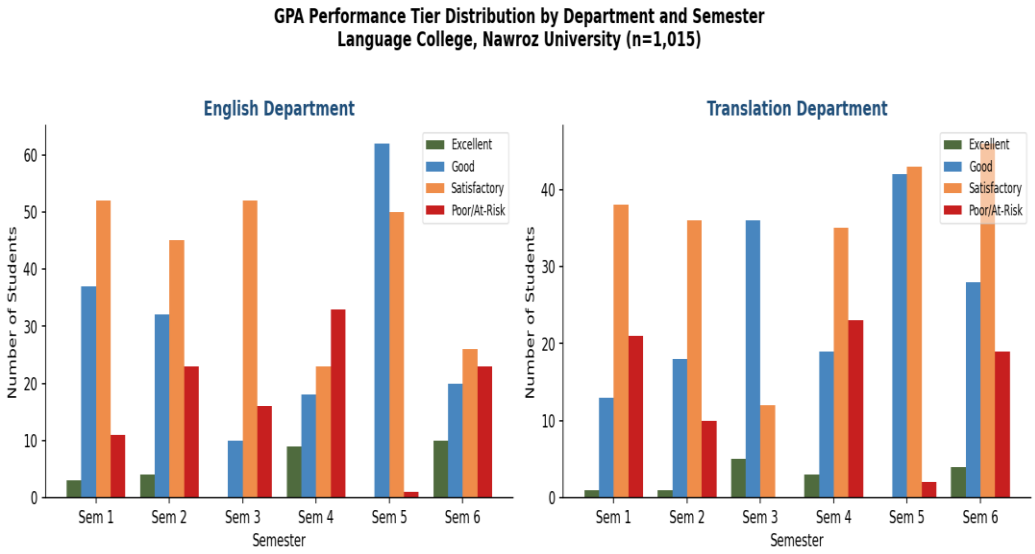

Figure 1. GPA Performance Tier Distribution by Department and Semester. English Department (left) and Translation Department (right), Language College, Nawroz University (n=1,015). Blue bars represent the Excellent tier, orange represents Good, green represents Satisfactory, and red represents Poor/At-Risk.

Table 3. Full Benchmarking Results — Six ML Classifiers on the Nawroz University Language College Dataset (n=1,015)

Algorithm	CV Accuracy	Test Accuracy	CV Precision	CV Recall	CV F1-Score
Random Forest	85.91%	92.12%	89.27%	80.78%	84.11%
KNN	88.28%	90.15%	90.28%	86.78%	88.06%
SVM (RBF) *	94.09%	96.06%	94.85%	91.24%	92.55%
Gradient Boosting	86.40%	88.67%	87.50%	80.20%	82.74%
Naive Bayes	82.96%	85.71%	87.78%	77.45%	80.59%
Decision Tree	67.49%	75.37%	69.93%	65.45%	66.66%

* Best overall model: SVM (RBF). CV = Stratified 5-Fold Cross-Validation. Test = independent 20% held-out partition.

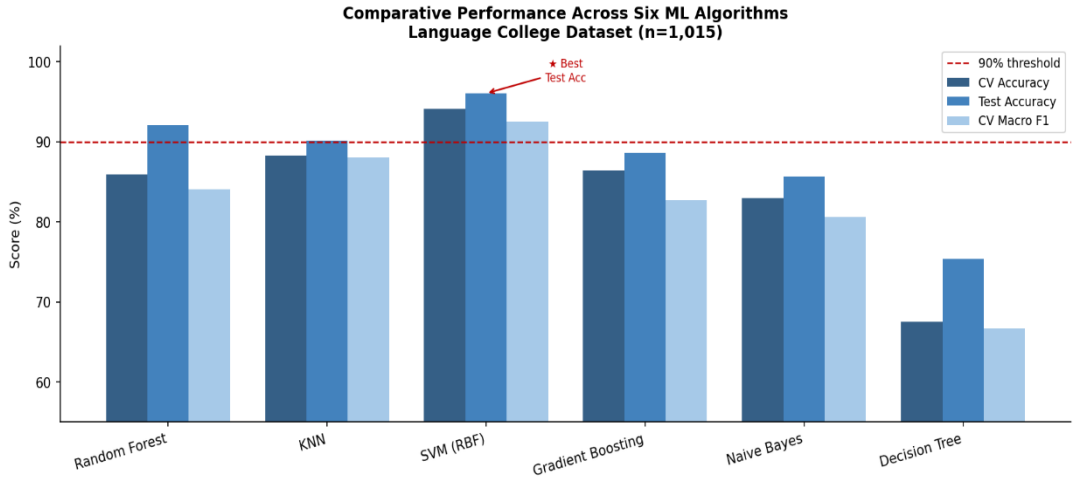

Figure 2. Comparative Performance Across Six ML Algorithms — CV Accuracy, Test Accuracy, and CV Macro F1-Score. Dashed red line marks the 90% threshold. Blue bars represent CV Accuracy, orange bars represent Test Accuracy, and green bars represent CV Macro F1-Score.

Figure 3 presents the confusion matrix for the Random Forest classifier, aggregated across all five cross-validation folds. The model demonstrated strong performance in classifying the Satisfactory tier, with most misclassifications occurring between adjacent performance tiers.

Importantly, no Poor/At-Risk students were misclassified into the Good or Excellent tiers, indicating that the model is effective in identifying academically vulnerable students.

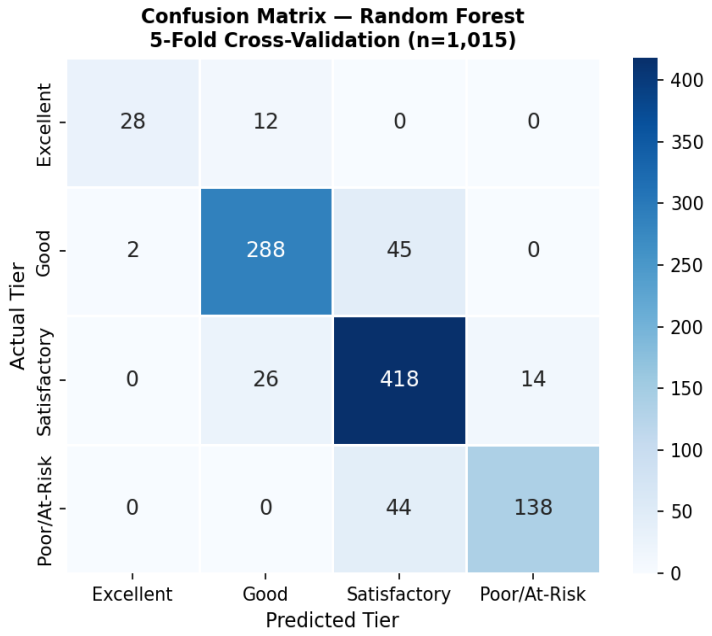

Figure 3. Confusion Matrix for the Random Forest Classifier aggregated across 5-fold cross-validation (n=1,015). Values on the diagonal represent correct tier assignments.

Figure 4 and Table 4 present the feature importance scores derived from the trained Random Forest model using the Gini impurity measure. Across the combined dataset from both departments, the importance values are relatively evenly distributed. Among the features, C5 exhibits the highest importance (0.193), followed by C4 (0.178), C1 (0.172), C3 (0.161), C2 (0.158), and C6 (0.138).

The three most influential features (C5, C4, and C1) together account for 54.3% of the total importance, indicating their dominant contribution to the model's predictive performance.

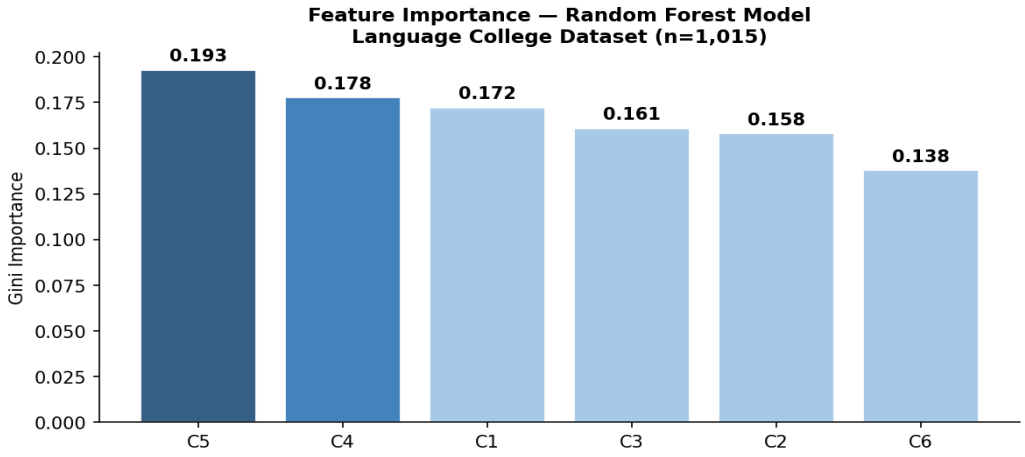

Figure 4. Gini-Based Feature Importance Scores — Random Forest Model. Feature importance is balanced across C1–C6, with C5 leading (0.193). Blue bars represent individual feature importance values.

Table 4. Course Feature Importance Ranking — Random Forest

Rank	Feature Code	Importance	Cumulative %
1	C5	0.193	19.3%
2	C4	0.178	37.1%
3	C1	0.172	54.3%
4	C3	0.161	70.4%
5	C2	0.158	86.2%
6	C6	0.138	100.0%

The SVM model with an RBF kernel achieved a test accuracy of 96.06% on an independent held-out set of 203 students, providing strong evidence that six course-level examination marks contain substantial predictive information. This high level of performance can be attributed to the ability of kernel-based methods to capture nonlinear relationships among course marks, allowing for more accurate separation of the four GPA performance tiers.

Random Forest represents a strong alternative, particularly due to its interpretability. The availability of feature importance scores offers clear practical value, especially when communicating findings to academic departments and decision-makers.

The relatively balanced distribution of feature importance (C5 = 0.193, C4 = 0.178, C1 = 0.172) suggests that no single course dominates the prediction process across the combined two-department dataset. This finding contrasts with the earlier single-department pilot, in which C6 (Poetry) exhibited a notably higher importance (0.307).

From a practical perspective, these results imply that effective early warning systems should consider all six course marks collectively, rather than relying on a single high-weight feature.

The framework's reliance solely on final examination marks provides a low-barrier pathway for practical implementation. A typical deployment scenario could proceed as follows: at the end of each examination marking period, the University Management System (UMS) extracts a student-by-course mark matrix, which is then input into the pre-trained SVM (RBF) classifier. The model assigns each student to a predicted performance tier, and those classified as Poor/At-Risk trigger automated advisory alerts to academic advisors and department heads.

This prediction-to-alert pipeline can be implemented without requiring changes to existing data collection practices, additional student surveys, or integration with external learning management or clickstream systems [15].

The conclusions of this study should be interpreted in light of several limitations. First, the findings are based on data from two departments within a single college at one institution; therefore, validation across additional colleges and institutions is necessary before broader generalization can be made. Second, the feature set is limited to final examination marks and does not include other potentially informative variables—such as attendance, coursework and mid-term grades, or students' socioeconomic background—that have been shown in prior research to influence predictive performance.

Third, the cross-sectional nature of the dataset prevented longitudinal tracking of individual students over multiple academic years. Finally, the feature importance rankings are likely to be curriculum-dependent and may vary across departments; as such, they should be re-evaluated when applying the framework in different academic contexts.

D. Conclusion

This paper presented, implemented, and evaluated a multi-algorithm machine learning framework for the early classification of student academic performance. The framework was applied to real examination data collected from the English and Translation Departments within the College of Languages at Nawroz University, Kurdistan Region of Iraq, covering all six semesters of the undergraduate program.

Using only six course-level final marks per student, the proposed approach successfully predicted four GPA performance tiers aligned with the Bologna Process across a dataset of 1,015 students. Among the six evaluated classifiers, the SVM with an RBF kernel achieved the best overall performance, with a held-out test accuracy of 96.06%, cross-validated accuracy of 94.09%, macro F1-score of 92.55%, macro precision of 94.85%, and macro recall of 91.24%. Random Forest and KNN also demonstrated strong performance.

Feature importance analysis indicated a relatively balanced contribution across courses, with C5 (0.193), C4 (0.178), and C1 (0.172) emerging as the most influential predictors, while no single course dominated the prediction process.

From an institutional perspective, one of the most significant implications of this study is that prediction accuracies exceeding 94% can be achieved without the need for additional data infrastructure. Specifically, the framework operates effectively without requiring LMS integration, student surveys, or clickstream data. This simplicity makes it well-suited for immediate deployment as an early warning component embedded within a University Management System (UMS).

Future work will focus on extending the dataset to include all departments across Nawroz University's colleges, enabling broader validation of the framework. In addition, incorporating longitudinal cohort tracking over multiple academic years and enriching the feature set with variables such as attendance and coursework grades may further enhance predictive performance.

E. Acknowledgment

The author would like to express sincere gratitude to the Department of Computer Science, College of Science, Nawroz University, for providing access to the academic records used in this study. All data were obtained for research purposes and were fully anonymized prior to analysis to ensure the protection of student privacy and confidentiality. The study was conducted in accordance with the institution's research ethics guidelines.

F. References

[1] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," Telematics and Informatics, vol. 37, pp. 13-49, 2019.

[2] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," WIREs Data Mining Knowl. Discov., vol. 10, no. 3, e1355, 2020.

[3] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," Procedia Comput. Sci., vol. 72, pp. 414-422, 2015.

[4] E. Alyahyan and D. Dustegor, "Predicting academic success in higher education: Literature review and best practices," Int. J. Educ. Technol. Higher Educ., vol. 17, Article 3, 2020.

[5] Y. A. Alsariera et al., "Assessment and evaluation of different machine learning algorithms for predicting student performance," Comput. Intell. Neurosci., 2022, Art. no. 4151487.

[6] Z. Xu, H. Yuan, and Q. Liu, "A comparative study on student performance prediction using machine learning," Educ. Inf. Technol., vol. 28, pp. 12039-12057, 2023.

[7] A. Villar and C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study," Discov. Artif. Intell., vol. 4, no. 2, 2024.

[8] H. Waheed et al., "Predicting academic performance of students from VLE big data using deep learning models," Comput. Human Behav., vol. 104, Art. no. 106189, 2020.

[9] A. Hellas et al., "Predicting academic performance: A systematic literature review," Proc. Companion ITiCSE 2018, pp. 175-199, 2018.

[10] W. Ahmed, M. K. Hasan, and G. Jeon, "Machine learning-based academic performance prediction with explainability," Sci. Rep., vol. 15, Art. no. 12353, 2025.
[11] J. Kruger, A. Britto, and J. P. Barddal, "An explainable machine learning approach for student dropout prediction," Expert Syst. Appl., vol. 233, Art. no. 120933, 2023.
[12] D. Baneres et al., "A predictive analytics infrastructure to support a trustworthy early warning system," Appl. Sci., vol. 11, no. 13, Art. no. 5781, 2021.
[13] K. E. Arnold and M. D. Pistilli, "Course signals at Purdue: Using learning analytics to increase student success," Proc. LAK 2012, pp. 267-270, 2012.
[14] Y. Wang, "Artificial intelligence in student management systems to enhance academic performance monitoring," Sci. Rep., vol. 15, Art. no. 35122, 2025.
[15] K. Alalawi, P. Santos, and K. Yacef, "An extended learning analytics framework integrating machine learning and pedagogical approaches," Int. J. Artif. Intell. Educ., 2024.
[16] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques," Appl. Sci., vol. 11, no. 1, Art. no. 237, 2021.