

Student Academic Performance Prediction Using Machine Learning: A Comparative Study of Decision Tree, Random Forest, and K-Nearest Neighbor

Isa Anita^{1*}, Restu Ramadhan Syafiani², Tarisya Alifiya Zahra³

¹²³Universitas Teknologi Digital, Indonesia

*Correspondent email: isa20125004@digitechuniversity.ac.id

Received: 11 Juni 2026 **Revised:** 3 Juli 2026 **Accepted:** 15 Oktober 2026 **Published:** 16 Oktober 2026

©2026 by the authors. Licensee Applied Engineering: Jurnal Rekayasa dan Sains Terapan
This article is an open access article distributed under the terms and conditions of the Creative Commons
Attribution (CC BY NC) license <https://creativecommons.org/licenses/by-nc/4.0/>.

Abstract. Forecasting how well students will perform has become a major theme in educational data mining (EDM), because spotting learners who are likely to struggle gives institutions time to step in before difficulties become entrenched. This narrative review draws together evidence from 17 peer-reviewed articles published between 2021 and 2026 that evaluated Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbor (KNN) classifiers on a variety of educational datasets. Where a clear ranking was possible, RF finished first or joint-first more often than either rival, a result the reviewed authors typically attribute to the way averaging many trees curbs overfitting and captures non-linear interactions among features. DT, however, matched or exceeded RF when the features were clean and strongly predictive, and it remains the easiest of the three to explain, whereas KNN swung widely with data quality, class balance, and feature scaling. Three methodological gaps recur across the corpus: models are rarely validated on data from other institutions, fairness across student groups is seldom examined, and the tension between accuracy and interpretability remains unresolved.

Key words: Educational Data Mining, Machine Learning, Random Forest, Decision Tree, K-Nearest Neighbor

Abstrak. Prediksi capaian akademik siswa kini menjadi tema utama dalam penambangan data pendidikan (*Educational Data Mining*/EDM), karena pengenalan dini siswa yang berpotensi mengalami kesulitan memberi waktu bagi institusi untuk bertindak sebelum masalahnya mengakar. Tinjauan naratif ini menghimpun bukti dari 17 artikel yang telah melalui proses *peer-review* dan terbit pada 2021–2026, yang menguji klasifikator *Decision Tree* (DT), *Random Forest* (RF), dan *K-Nearest Neighbor* (KNN) pada beragam dataset pendidikan. Pada studi yang peringkatnya dapat ditentukan secara jelas, RF lebih sering berada di posisi pertama atau pertama bersama dibandingkan dua algoritma lainnya; para penulis studi umumnya mengaitkan hal ini dengan mekanisme rata-rata banyak pohon yang menekan *overfitting* dan menangkap interaksi fitur yang tidak linear. Namun, DT menyamai atau melampaui RF ketika fitur bersih dan sangat prediktif, serta tetap paling mudah dijelaskan, sedangkan kinerja KNN berayun lebar bergantung pada kualitas data, keseimbangan kelas, dan penskalaan fitur. Tiga kesenjangan metodologis muncul berulang: model jarang divalidasi pada data institusi lain, keadilan antarkelompok siswa jarang dikaji, dan pertentangan antara akurasi dan interpretabilitas belum terselesaikan.

Kata Kunci: Penambangan Data Pendidikan, Pembelajaran Mesin, *Random Forest*, Pohon Keputusan, *K-Nearest Neighbor*

1. Introduction

Grades and other achievement indicators are widely treated as a gauge of how well a student is doing and, by extension, of the quality of an institution's teaching. Spotting students who are likely to fall behind is nonetheless difficult, since conventional monitoring tends to react after results have already declined instead of warning in advance. At the same time, the move of teaching and administration onto digital platforms has generated extensive student records, such as attendance logs, test results, behavioral traces, and demographic profiles, that can inform evidence-based decisions (Yağcı, 2022). Educational data mining (EDM) applies computational methods to uncover useful regularities in such records, and it has become a promising route toward flagging struggling students early and with reasonable accuracy (Paul et al., 2025).

Machine learning (ML) sits at the core of much EDM work because models can infer patterns directly from past records rather than being hand-coded for every case (Khairy et al., 2024). Three algorithms recur especially often in this literature. Decision Tree (DT) classifiers are popular because the splitting rules they learn can be read and explained to teachers and administrators. Random Forest (RF) builds many trees on resampled data and combines their votes, which usually yields better generalization and greater tolerance to overfitting than a single tree. K-Nearest Neighbor (KNN) is simple to implement; being non-parametric, it can follow different data distributions provided that the features are suitably prepared and scaled (Dahal & Shakya, 2022).

Even though this body of work keeps growing, it leaves several questions open. To begin with, studies that pit the algorithms against one another disagree about which one wins, and the disagreement can usually be traced to differences in the data, in how features were chosen, and in the metrics used. In addition, a good share of the work draws on a single institution, which makes it unclear whether the findings would travel elsewhere. A third issue is the balance between predictive power and transparency, a matter of real importance if such models are to be adopted in schools and universities, yet one that has received limited attention (Hussain, 2025; Ifeanyi Martins Nwanegbo et al., 2025). A review that brings these results together across settings is therefore warranted.

The present article is a narrative literature review of 17 peer-reviewed studies, published from 2021 to 2026, in which DT, RF, and KNN were compared for

predicting student academic outcomes. It pursues three aims: (1) to collate the accuracy figures and main conclusions reported in each study; (2) to identify where the studies agree and where they diverge on how the algorithms rank; and (3) to point out methodological problems that remain unsolved and suggest directions for later work. By bringing recent evidence together, the review is intended to help researchers and practitioners choose and deploy suitable models.

2. Research Methodology

2.1 Research Design

A narrative review design was adopted. It was preferred because it allows the evidence on the effectiveness of DT, RF, and KNN to be interpreted and critiqued in a flexible way. In contrast to a systematic review, which follows a rigid protocol, a narrative review can range over many kinds of educational data and bring out research gaps along the way.

2.2 Search Strategy

Candidate articles were first located with Consensus, an AI-assisted search engine for scholarly literature. Queries combined terms such as "student academic performance prediction", "machine learning algorithms", "Decision Tree vs Random Forest vs K-Nearest Neighbor", and "educational data mining". The search covered publications from 2021 to 2026 so that only recent methods would be considered. To widen coverage, additional keyword searches were run on publisher platforms (Frontiers, Nature Portfolio, MDPI, Wiley, and Baghdad Science Journal), and the figures reported for each added article were checked against the article itself.

2.3 Selection Criteria

The following inclusion and exclusion criteria were used to safeguard the relevance and quality of the reviewed articles:

1. Inclusion Criteria

- Peer-reviewed journal articles or other scholarly publications issued between 2021 and 2026.
- Articles that predict student academic performance with ML.

- Articles comparing at least two of the three target algorithms (DT, RF, or KNN).
- Articles reporting quantitative performance metrics (e.g., accuracy, F1-score, R^2).

2. Exclusion Criteria

- Articles without empirical results, and editorials or opinion pieces.
- Studies set outside the educational context

3. Result and Discussion

3.1 Result

To provide a comparative overview of machine learning algorithms in predicting students' academic performance, Table 1 summarizes the findings of previous studies employing Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbor (KNN). The comparison highlights differences in dataset characteristics, classification accuracy, and the best-performing algorithms across the studies. These findings provide a basis for evaluating the relative performance of the three algorithms and identifying their strengths and limitations in educational data classification.

Table 1. Comparison Table

Author (Year)	Dataset Description	Decision Tree Accuracy (%)	Random Forest Accuracy (%)	K-Nearest Neighbor Accuracy (%)	Best Algorithm	Findings
Holicza and Kiss (2023)	Online & offline GPA data	100,00	95,00-100,00	70,00-100,00	Decision Tree (offline); all three tied at 100% (online)	On the offline data DT led with 100%, against 95% for RF and 70% for KNN. On the online data all three reached 100%, a result to be read with caution because perfect scores may signal overfitting.
Gul et al. (2025)	Demographic & academic	82,42	86,21	79,07	CatBoost (RF best among	RF scored 86.21%, ahead of DT (82.42%) and KNN

	(1,000 records)				the three target algorithms)	(79.07%), but CatBoost was the best of the eight models tested (87.46%).
Dahal and Shakya (2022)	Higher secondary students (two datasets of about 1,000 records)	85,99-100,00	88,85-100,00	81,85-100,00	CART (Decision Tree)	The authors judged the CART decision tree the strongest classifier of student results, especially on small datasets.
Nabil et al. (2021)	University course grades	84,00-87,00	87,00-91,00	74-91%	Deep Neural Network (DNN)	With oversampling (SMOTE), a Deep Neural Network obtained the best F1-score and the lowest error, ahead of the conventional classifiers.
Khairy et al. (2024)	Undergraduate statistics course	98,70	98,70	89,60	Random Forest & Decision Tree	RF and DT tied at 98.7% and showed no overfitting between the training and testing phases; KNN trailed at 89.6%.
Hussain (2025)	Student records from Iraqi university	97,00	98,00	99,00	Logistic Regression (KNN best among the three target algorithms)	Logistic Regression scored highest (99.50%), followed by KNN and SVM at 99%. DT was slightly lower but remains useful for showing how features interact.
Maulana et al. (2023)	395 students from two Portuguese secondary schools	$R^2 = 0,88$	$R^2 = 0,91$	$R^2 = 0,84$	Random Forest	Regression metrics were used instead of accuracy. RF obtained the highest R^2 (0.91) and the lowest RMSE (1.44).
Bujang et al. (2021)	Multi-institutional student grade data.	99,1	99,5	99,3	Random Forest	RF obtained the best F-measure, and SMOTE helped balance the proportion of minority-class instances.
Yağcı (2022)	Turkish Language	Not Evaluate	74,60	69,90	Random Forest &	RF and a neural network tied for the

	-I exam records, 2019–2020 (1,854 students)	d			Neural Network	highest accuracy (74.6%) and also had the two highest AUC values; KNN reached 69.9%.
Vives et al. (2024)	Data consisting of 661 rows of computer programming students.	93,40-100,00	92,20-100,00	88,40	Deep Neural Network & Long Short-Term Memory	DT and RF tended toward false-positive bias when the class distribution was unstable; DNN and LSTM combined with SMOTE and GAN gave much more stable accuracy.
Nwanegbo et al. (2025)	The Student Performance dataset on demographics, family, and grades.	Not Evaluated	86,00	75,00-79,00	Support Vector Classifier & Random Forest	SVC and RF were the strongest models, while KNN's results dropped under class imbalance.
Adeyanju et al. (2025)	The Student Performance dataset on demographics, family, and grades.	76,31	75,88	72,40	Decision Tree	After tuning, DT scored highest (76.31%), narrowly ahead of RF and KNN, with statistically significant differences; the authors stressed its efficiency and transparency.
Ahmed (2024)	32,005 student records from Wollo University and Kombolcha Institute of Technology, Ethiopia (2017–2022)	90,99-93,41	Not Evaluated	85,38-87,38	SVM (96,03); DT best among target algorithms	Repeated 10-fold cross-validation with tuning raised every model. SVM led, DT came second, KNN third, and Naive Bayes was lowest (77,38 before tuning, 83,32 after).

Alsumaidaie et al. (2024)	30,000 student records (24,000 training, 6,000 testing)	Not Evaluated	94,03	91,65	Extra Trees (98,15); RF best among target algorithms	Extra Trees topped the three models tested, RF came second, and KNN third.
Paul et al. (2025)	33,627 student records from colleges affiliated with Dibrugarh University, India	94,23-97,89	93,98-95,93	90,23-90,91	Logistic Regression (98,43); DT best among target algorithms	DT beat RF and KNN under every splitting criterion, though logistic regression, deep learning, and gradient boosting edged ahead overall. SHAP showed credit totals and marks as the main drivers.
Agal (2026)	More than 100,000 synthetic university-intake records, 28 attributes	F1 0,67	F1 0,81	Not Evaluated	Ordinal Neural Network (F1 0,85)	RF clearly outperformed DT; LightGBM was the best standard classifier (F1 0,84). Class 12 marks were the dominant predictor.
Ewais et al. (2026)	1,419 records, 12 features, one mandatory course at a Palestinian university	$R^2 = 0,86$	$R^2 = 0,93$	Not Evaluated	Random Forest	RF regression gave the highest R^2 (0.93) and lowest MAE (0.05), significantly ahead of the five other regressors tested, including DT.

Note. Values are accuracy (%) unless another metric is stated (F1 = F1-score; R^2 = coefficient of determination). "Not Evaluated" means the algorithm was not part of the study; "Not Specified" means no value was reported.

3.2 Discussion

Table 1 shows that the ranking of the three algorithms depends on the educational setting. For 14 of the 17 studies, a ranking among the target algorithms could be read directly from the reported figures. RF finished first or joint-first in nine of these (Agal, 2026; Bujang et al., 2021; Ewais et al., 2026; Gul et al., 2025; Ifeanyi Martins Nwanegbo et al., 2025; Khairy et al., 2024; Maulana et al., 2023; S. Ibrahim Alsumaidaie

et al., 2024; Yağcı, 2022), DT in five (Adeyanju et al., 2025; Ahmed, 2024; Dahal & Shakya, 2022; Khairy et al., 2024; Paul et al., 2025), and KNN in one (Hussain, 2025); studies with ties are counted under every algorithm involved. The other three (Holicza & Kiss, 2023; Nabil et al., 2021; Vives et al., 2024) reported overlapping score ranges and were left unranked. RF's advantage is usually explained by its combination of many decorrelated trees, which dampens overfitting and can capture non-linear relations among features. The advantage is far from universal, however. DT led on relatively small datasets (Adeyanju et al., 2025; Dahal & Shakya, 2022) and also on two very large ones (Ahmed, 2024; Paul et al., 2025), which suggests that the cleanliness and predictive strength of the features matter more than sample size. Because a single tree is far easier to explain than a forest, DT's transparency may therefore cost little accuracy when conditions are favorable.

KNN showed the most erratic behavior. It produced the top result among the three algorithms in Hussain (2025), at 99%, and came a close second in Bujang et al. (2021). Yet it fell behind RF in Yağcı (2022), with 69.9% against 74.6%; in Nwanegbo et al. (2025), with 75–79% against 86%; in Alsumaidaie et al. (2024), with 91.65% against 94.03%; and in Paul et al. (2025), with about 90.2–90.9% against up to 95.9%. Because it relies on distances between records, KNN is vulnerable to differences in feature scale, to class imbalance, and to the choice of k , which is consistent with the broader view that careful preprocessing is a precondition for good KNN results.

The characteristics of the data emerged as a recurring influence on the outcome. Resampling with SMOTE, applied by Bujang et al. (2021) and Nabil et al. (2021), was reported to yield more balanced and dependable predictions, particularly for minority-class students at risk of failure. Including behavioral and demographic variables, such as attendance, family background, and socioeconomic indicators, was likewise linked to better accuracy irrespective of the algorithm (Holicza & Kiss, 2023; Maulana et al., 2023; Yağcı, 2022). More recent studies point the same way for different reasons: in a dataset of 33,627 students, marks and credit totals carried most of the predictive weight while gender contributed almost nothing (Paul et al., 2025), and prior school achievement dominated the predictions in the synthetic admissions data of Agal (2026). Feature engineering therefore appears to be at least as important as the choice of algorithm.

Several studies also found that deep learning models (DNN, LSTM) outperformed all three target algorithms, especially on sequential or high-dimensional data (Nabil et

al., 2021; Vives et al., 2024), and recent work shows a similar pattern for other model families: an ordinal neural network and LightGBM beat both RF and DT in Agal (2026), Extra Trees surpassed RF in Alsumaidaie et al. (2024), and logistic regression, deep learning, and gradient boosting edged past the best DT in Paul et al. (2025). Although these models fall outside the scope of this review, they hint at a ceiling for DT, RF, and KNN on complex educational data and point toward hybrid or more advanced models as a direction for future research. Ewais et al. (2026) are a counterexample: RF remained the strongest of the six regression models they tested.

3.3 Similarities Among Studies

Despite differences in context and method, the reviewed studies have several features in common. First, all but one drew on real student records such as grades, attendance, and family background; the exception, Agal (2026), used a public synthetic dataset modeled on university admissions. Second, each study compared at least two of the three target algorithms, which makes cross-algorithm comparison possible. Third, tree ensembles, RF in particular, usually did better than single classifiers, although exceptions exist (for example, DT outperformed RF in Paul et al., (2025)). Finally, most of the studies stressed that feature selection and preprocessing shape model performance, whichever algorithm is used.

3.4 Differences Among Studies

The studies also differ considerably in scale and method. Sample sizes range from a few hundred records, such as 395 students in Maulana et al. (2023), to more than 30,000 in Ahmed (2024), Alsumaidaie et al. (2024), and Paul et al. (2025), and over 100,000 synthetic records in Agal (2026). Some studies added deep learning models such as DNN and LSTM, which outperformed the conventional classifiers in their settings (Nabil et al., 2021; Vives et al., 2024), while others added boosted or hybrid models (Agal, 2026; Gul et al., 2025). Evaluation practices diverge as well: most studies reported accuracy, others relied on F1-score (Agal, 2026), and a few used regression metrics such as R^2 (Ewais et al., 2026; Maulana et al., 2023). Prediction targets also vary, from binary pass/fail labels to multi-class grade bands and continuous scores. Together these differences make direct comparison across studies difficult.

3.5 Research Gap

Despite the amount of research on this topic, several gaps remain. Most of the reviewed studies were conducted within a single institution, and some within a single course (Ewais et al., 2026). Even a dataset of more than 33,000 records came from one university system (Paul et al., 2025). This limits how far the findings can be generalized to other educational systems or cultures. Questions of fairness and bias, in particular whether predictions are equally reliable for different demographic groups, have received little attention in the reviewed literature. Finally, the trade-off between interpretability and predictive power remains unresolved: a single tree gives transparent rules that teachers can act on, while higher-scoring ensembles behave more like black boxes, which hampers practical adoption in schools and universities.

4. Conclusion

This narrative review examined 17 peer-reviewed studies that compared DT, RF, and KNN for predicting student academic performance. No algorithm won everywhere, but RF was the most dependable overall: among the 14 studies where a ranking could be drawn, it came first or joint-first in nine, compared with five for DT and one for KNN (ties are counted under each algorithm involved). Its edge is generally attributed to its ensemble structure, which curbs overfitting and handles complex feature interactions. DT matched or exceeded RF when the features were clean and strongly predictive, in small as well as very large datasets, and it keeps a practical advantage in educational settings because its decision rules are easy to interpret. KNN was the least consistent: strong in some studies (Hussain, 2025) but clearly behind in others (Yağcı, 2022), which confirms its dependence on data quality and preprocessing.

Beyond the choice of algorithm, the review shows that dataset characteristics, feature engineering, and preprocessing, resampling methods such as SMOTE among them, are equally decisive for model performance. Practitioners building student performance prediction systems should therefore devote as much effort to data preparation and feature curation as to algorithm selection. Because the evidence still comes largely from single institutions, future work should test these models across institutions and report fairness and interpretability alongside accuracy.

References

- Adeyanju, T. M., Aziz, A. A., & Safei, S. (2025). A Comparative Analysis of Classification Algorithms on Student Academic Performance. *Semarak International Journal of Machine Learning*. <https://api.semanticscholar.org/CorpusID:277254252>
- Agal, S. (2026). A machine learning framework with a public synthetic data set for early student profiling and outcome prediction. *Scientific Reports*, 16(1), 25730. <https://doi.org/10.1038/s41598-026-55514-9>
- Ahmed, E. (2024). Student Performance Prediction Using Machine Learning Algorithms. *Applied Computational Intelligence and Soft Computing*, 2024(1), 4067721. <https://doi.org/10.1155/2024/4067721>
- Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. Md. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. *IEEE Access*, 9, 95608–95621. <https://doi.org/10.1109/ACCESS.2021.3093563>
- Dahal, N. P., & Shakya, S. (2022). A Comparative Analysis of Prediction of Student Results Using Decision Trees and Random Forest. *Journal of Trends in Computer Science and Smart Technology*, 4(3), 113–125. <https://doi.org/10.36548/jtcsst.2022.3.001>
- Ewais, Ahmed., Arram, A., Diab, S., Alkhazaleh, H. A., Himeur, Y., & Ashqar, H. I. (2026). Using machine learning for early identification of At-risk students. *Scientific Reports*, 16(1), 19944. <https://doi.org/10.1038/s41598-026-48426-1>
- Gul, M. N., Abbasi, W., Babar, M. Z., Aljohani, A., & Arif, M. (2025). Data driven decisions in education using a comprehensive machine learning framework for student performance prediction. *Discover Computing*, 28(1), 153. <https://doi.org/10.1007/s10791-025-09585-3>
- Holicza, B., & Kiss, A. (2023). Predicting and Comparing Students' Online and Offline Academic Performance Using Machine Learning Algorithms. *Behavioral Sciences*, 13(4), 289. <https://doi.org/10.3390/bs13040289>
- Hussain, S. M. (2025). Predicting Student Performance Using Data Mining and Machine Learning Techniques. *JOURNAL OF UNIVERSITY OF BABYLON for Pure and Applied Sciences*, 19–36. <https://doi.org/10.29196/jubpas.v33i1.5640>
- Ifeanyi Martins Nwanegbo, Costa Denison-George, Daniel Jonathan, & Nonso Fredrick Chiobi. (2025). PREDICTING STUDENT ACADEMIC PERFORMANCE USING MACHINE LEARNING. *International Journal of Modeling and Applied Science Research*. <https://doi.org/10.70382/caijmasr.v8i9.015>
- Khairy, D., Alharbi, N., Amasha, M. A., Areed, M. F., Alkhalaf, S., & Abougalala, R. A. (2024). Prediction of student exam performance using data mining classification algorithms. *Education and Information Technologies*, 29(16), 21621–21645. <https://doi.org/10.1007/s10639-024-12619-w>
- Maulana, A., Idroes, G. M., Kemala, P., Maulydia, N. B., Sasmita, N. R., Tallei, T. E., Sofyan, H., & Rusyana, A. (2023). Leveraging Artificial Intelligence to Predict Student

- Performance: A Comparative Machine Learning Approach. *Journal of Educational Management and Learning*, 1(2), 64–70. <https://doi.org/10.60084/jeml.v1i2.132>
- Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access*, 9, 140731–140746. <https://doi.org/10.1109/ACCESS.2021.3119596>
- Paul, R., Sarker, S., El Aouifi, H., Hussain, S., Baruah, A. K., & Gaftandzhieva, S. (2025). Analyzing dropout of students and an explainable prediction of academic performance utilizing artificial intelligence techniques. *Frontiers in Education*, 10, 1698505. <https://doi.org/10.3389/feduc.2025.1698505>
- S. Ibrahim Alsumaidaie, M., Adil Nafea, A., Abbas Mukhlif, A., D. Jalal, R., & M AL-Ani, M. (2024). Intelligent System for Student Performance Prediction Using Machine Learning. *Baghdad Science Journal*, 21(12), 3877–3891. <https://doi.org/10.21123/bsj.2024.9643>
- Vives, L., Cabezas, I., Vives, J. C., Reyes, N. G., Aquino, J., Córdor, J. B., & Altamirano, S. F. S. (2024). Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks. *IEEE Access*, 12, 5882–5898. <https://doi.org/10.1109/ACCESS.2024.3350169>
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>