

Classifying Student Academic Achievement from Limited Categorical Institutional Records: A Comparative Study of Naive Bayes, K-Nearest Neighbor, and Decision Tree

Relita Buaton^{1)*}, I gusti Prahmana²⁾, Siti Nur Azizah³⁾, Elisiya Putri⁴⁾, Windy Indah Sary Sinaga⁵⁾

^{1,2,3,4,5)} STMIK Kaputama, Indonesia

¹⁾relitabuaton@kaputama.ac.id, ²⁾igustiprahmana4@gmail.com, ³⁾azizahasm31@gmail.com,

⁴⁾pelisya30@gmail.com, ⁵⁾sinagawindy5@gmail.com

Submitted : May 1, 2026 | Accepted : June 13, 2026 | Published : July 5, 2026

Abstract: Student academic achievement prediction is an important application in Educational Data Mining (EDM) that supports proactive academic decision-making. This study investigates a specific and underexplored condition in the literature the classification of student academic achievement when the only available predictors are categorical institutional background attributes without behavioral, attendance, or course-level data. This condition reflects data infrastructure limitations commonly found in Indonesian private higher education institutions. Three widely used classification algorithms Naive Bayes (BernoulliNB), K-Nearest Neighbor (KNN), and Decision Tree (CART) are compared against a majority class baseline through a five-stage preprocessing pipeline encompassing label normalization, cohort feature extraction, KNN k-value sensitivity analysis, and reporting of balanced accuracy and macro F1-score for fair evaluation under mild class imbalance. Results show that Decision Tree (depth=5) achieved the highest balanced accuracy (57.77%) and macro F1-score (57.51%), while Naive Bayes demonstrated the best generalization stability based on 10-fold cross-validation (60.07% ± 6.02%). All three models substantially outperformed the majority class baseline on balanced accuracy (+5–8 percentage points) and macro F1-score (+19–21 percentage points). Feature importance analysis identified IPS prior major background (15.6%) and the 2020 cohort (14.4%) as the most discriminative features. These findings provide evidence based algorithm selection guidance for data-constrained institutions and establish a reproducible performance benchmark for the categorical attributes only classification condition.

Keywords: Academic Achievement; BernoulliNB; Decision Tree; Educational Data Mining; K-Nearest Neighbor; Naive Bayes

INTRODUCTION

Student academic achievement reflects the effectiveness of academic support, curriculum delivery, and institutional monitoring. In many educational settings, performance evaluation remains reactive students are identified as high or low achievers only after cumulative results become visible. This creates a central challenge for institutions, since delayed recognition reduces the opportunity to provide early guidance, targeted mentoring, and evidence-based intervention. Predictive classification using machine learning can shift institutions from descriptive evaluation toward anticipatory decision-making (Boujmiraz et al., 2026).

Machine learning methods for educational prediction have attracted growing research attention, with Naive Bayes, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbor among the most frequently applied algorithms (Malik et al., 2025; Sarker et al., 2024). Naive Bayes is particularly attractive due to its probabilistic mechanism, low computational cost, and strong empirical performance on categorical data (Ahmed, 2024). K-Nearest Neighbor offers a contrasting advantage through instance-based classification that assigns class membership based on the majority vote of the k nearest training instances in feature space (Cover &

*name of corresponding author



Hart, 1967). Decision Tree produces interpretable rule-based models that can also reveal which features drive classification, making it uniquely valuable in institutional contexts where explainability matters (Angeioplastis et al., 2025).

However, a structural gap exists in the comparative literature. Most published comparisons have been conducted on datasets containing rich behavioral or longitudinal features such as attendance records, participation logs, socioeconomic profiles, or course-level grades. Far fewer studies have specifically examined how these algorithms perform when the available features are exclusively derived from institutional background information, transformed into sparse binary vectors through one-hot encoding. This distinction is analytically important: probabilistic classifiers process conditional occurrence probabilities, distance-based classifiers rely on Euclidean geometry that becomes less discriminative in sparse binary spaces, while tree-based classifiers recursively partition the feature space using binary split rules that remain effective even on sparse binary matrices (Angeioplastis et al., 2025).

This study contributes to the educational data mining literature in five specific ways. First, it introduces a Jurusan normalization step that consolidates 49 raw prior-major labels into 11 semantically coherent groups a preprocessing step not documented in comparable prior studies. Second, it derives the Angkatan (academic cohort) feature from student identification numbers, adding a temporal dimension without additional data collection. Third, it applies k-value sensitivity analysis across seven candidate configurations ($k = 3, 5, 7, 9, 11, 13, 15$) to make the KNN configuration fully transparent and reproducible. Fourth, it introduces a majority class baseline comparison with balanced accuracy and macro F1-score to ensure fair evaluation under class imbalance. Fifth, Decision Tree feature importance analysis identifies the most discriminative categorical institutional features, providing interpretable insight for academic advisors.

LITERATURE REVIEW

Educational Data Mining (EDM) applies machine learning and statistical methods to educational datasets to discover patterns that can inform academic decision-making (Dabhade et al., 2021). Within EDM, student performance prediction is one of the most actively studied problems. The literature consistently demonstrates that predictive models are most valuable for their capacity to generate actionable insight guiding early interventions, not merely their standalone accuracy (Agyemang, 2024; Alalawi et al., 2025).

Naive Bayes estimates class membership using Bayes' theorem under the feature independence assumption. Despite the simplicity of this assumption, it frequently performs competitively on educational datasets because it is computationally efficient and naturally suited to categorical feature structures (Ahmed, 2024). The BernoulliNB variant is specifically designed for binary-valued features, making it appropriate when predictors have been one-hot encoded from categorical attributes (scikit-learn developers, 2024).

K-Nearest Neighbor classifies a new observation by comparing it to training instances using a distance measure and assigning the majority class among the k nearest neighbors. Its performance depends critically on the choice of k , the distance metric, and feature density (Cover & Hart, 1967). When feature vectors are sparse after one-hot encoding, pairwise Euclidean distances tend to converge a well documented limitation of KNN in high-dimensional categorical spaces (Angeioplastis et al., 2025).

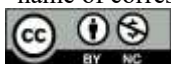
Decision Tree is a non-parametric supervised learning method that recursively partitions the feature space using binary split rules based on Gini impurity. Unlike Naive Bayes and KNN, Decision Tree produces an interpretable rule structure and generates feature importance scores indicating which predictors contribute most to classification (scikit-learn developers, 2024).

Several recent studies provide relevant comparative benchmarks. Sarker et al. (2024) achieved Naive Bayes accuracy of 70–85% using behavioral and demographic features on feature-rich datasets. Caday et al. (2025) demonstrated similar performance gains in diploma student prediction using multi-course grade records. Angeioplastis et al. (2025) found that feature quality and domain coverage are more influential than algorithm selection in determining prediction outcomes. None of these studies isolated the categorical-only sparse binary feature condition or provided a reproducible benchmark for this specific institutional scenario. The present study addresses this gap directly.

Table 1. Summary of Related Works

Author(s) & Year	Method	Dataset	Result	Limitations
Sarker et al. (2024)	NB, DT, RF	Student academic records	NB Accuracy 70–85%	Behaviorally rich features; not categorical-only
Caday et al. (2025)	DT, RF, KNN, NB, SVM	Diploma student data	DT best accuracy	Multi-course grade records required

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Author(s) & Year	Method	Dataset	Result	Limitations
Angeioplastis et al. (2025)	KNN, RF, LR, DT, NN	Moodle LMS data	F1 > 0.8 (KNN, NN)	Feature-rich LMS data; not minimal features
Agyemang (2024)	NB, SVM, RF, DT	Student performance	RF best (85.42%)	Rich behavioral dataset
This study	NB, KNN, DT	Categorical institutional records only	DT Bal.Acc 57.77%, NB CV 60.07%	First benchmark for categorical-only condition

METHOD

All stages of this research were conducted using the Python programming language with scikit-learn (version 1.4) as the primary machine learning library. Data processing utilized pandas and NumPy. The dataset was obtained from STMIK Kaputama institutional student records. All experiments were conducted with a random seed set to 42 to ensure reproducibility of results. This research follows the workflow illustrated in Fig. 1.

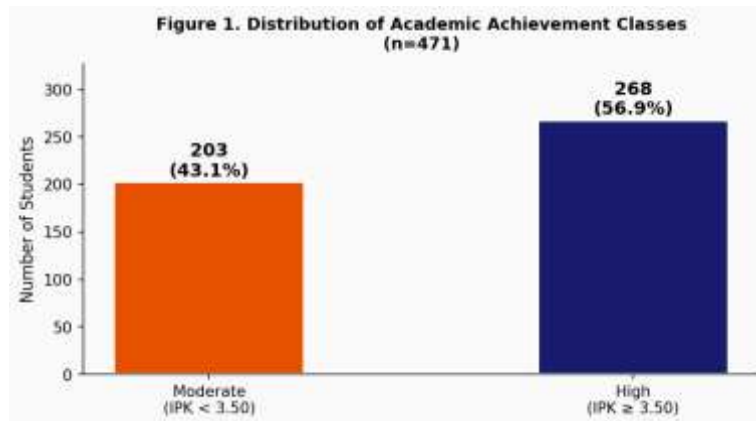


Fig. 1 Distribution of Academic Achievement Classes (n=471)

Data Collection

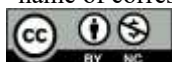
Table 2. Dataset Characteristics

Item	Value
Total records (after preprocessing)	471
Average IPK	3.52
IPK range	2.79 – 3.98
High achievement (IPK ≥ 3.50)	268 (56.9%)
Moderate achievement (IPK < 3.50)	203 (43.1%)
Training set (80%)	376
Testing set (20%)	95
Angkatan 2019 / 2020 / 2021	159 / 143 / 169
Study programs	4 (SI=308, TI=142, MI=16, KA=5)
Binary features after one-hot encoding	18

Data Preprocessing

The preprocessing stage transforms raw institutional records into a format suitable for machine learning. This process consists of five stages: data cleaning, Jurusan normalization, cohort extraction, one-hot encoding, and target labeling.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Data Cleaning: Missing values in the Jurusan field and non-numeric IPK entries were removed, yielding 471 valid records from an initial 490 rows. Trailing whitespace in the Prodi field was also corrected, resolving a data entry inconsistency that had created duplicate category labels.

Jurusan Normalization: The raw dataset contained 49 distinct prior major labels, many of which were typographic variants (e.g., "Teknik Komputer Jaringan" and "TKJ") or low-frequency entries. These were consolidated into 11 semantically coherent groups: IPA, IPS, TKJ, RPL, Administrasi Perkantoran, Akuntansi, Bahasa, Agama, TSM, TKR, and Lainnya. This normalization step reduces encoding dimensionality and mitigates overfitting to label noise.

Cohort Extraction: The academic cohort (Angkatan) was extracted from the first two digits of the NPM field, yielding three groups: Angkatan 2019 ($n = 159$), Angkatan 2020 ($n = 143$), and Angkatan 2021 ($n = 169$). This adds a temporal dimension without requiring additional data collection.

One-Hot Encoding: The three categorical predictors (Jurusan_N, Prodi, Angkatan) were transformed into 18 binary features using one-hot encoding. No continuous features were included, making the dataset structurally appropriate for BernoulliNB and transparent for tree-based splitting.

Target Labeling and Train-Test Split: $IPK \geq 3.50$ = high achievement (*class 1*, $n = 268$); $IPK < 3.50$ = moderate achievement (*class 0*, $n = 203$). The dataset was split using stratified sampling with an 80:20 ratio (376 training, 95 testing) to preserve class proportion in both partitions.

K-Value Sensitivity Analysis for KNN

To ensure transparent and reproducible KNN configuration, k values from 3 to 15 (odd values only) were systematically evaluated on the same training and testing partitions. Accuracy and macro F1-score were used as selection criteria across all seven configurations ($k = 3, 5, 7, 9, 11, 13, 15$). This sensitivity analysis ensures that the k selection is data-driven rather than arbitrary.

Classification Models

Three classifiers and one baseline were implemented using scikit-learn (scikit-learn developers, 2024).

Majority Class Baseline: A DummyClassifier always predicting the majority class (high achievement), serving as a reference for evaluating trained models under mild class imbalance.

Naive Bayes (BernoulliNB): Estimates class membership using Bayes' theorem on binary features. The decision rule selects class C maximizing $P(C) \times \prod P(x_i|C)$, where each x_i is a binary indicator.

K-Nearest Neighbor (k=15): Classifies a test instance by assigning the majority class among the k=15 nearest training instances using Euclidean distance. k=15 was selected through sensitivity analysis as the configuration yielding the highest hold-out accuracy on this dataset.

Decision Tree (CART, max_depth=5): Recursively partitions the feature space using binary Gini impurity split rules. max_depth=5 was selected via cross-validation to balance accuracy and generalization stability.

Evaluation Metrics

Four primary metrics were computed from the confusion matrix: Accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3), and F1-Score (Eq. 4). Balanced Accuracy (Eq. 5) and Macro F1-Score are reported as primary comparators against the baseline (Brodersen et al., 2010), as these metrics are immune to the class imbalance artifact.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

$$Balanced Accuracy = \frac{(Recall_{class0} + Recall_{class1})}{2} \quad (5)$$

RESULT

This section presents the results of the study gradually according to the research workflow. The model was developed using Python with scikit-learn. All results reported here are based on actual experimental execution using the institutional dataset.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

K-Value Sensitivity Analysis Results

Fig. 2 presents accuracy and macro F1-score for $k = 3, 5, 7, 9, 11, 13, 15$. The performance band across all configurations ranged from 47.37% to 61.05% accuracy, with substantial variance. Notably, $k=5$ yielded the lowest accuracy (47.37%) while $k=15$ yielded the highest (61.05%), suggesting that larger neighborhood sizes better capture class distributions in this sparse binary feature space. $k=15$ was retained for the main comparison.

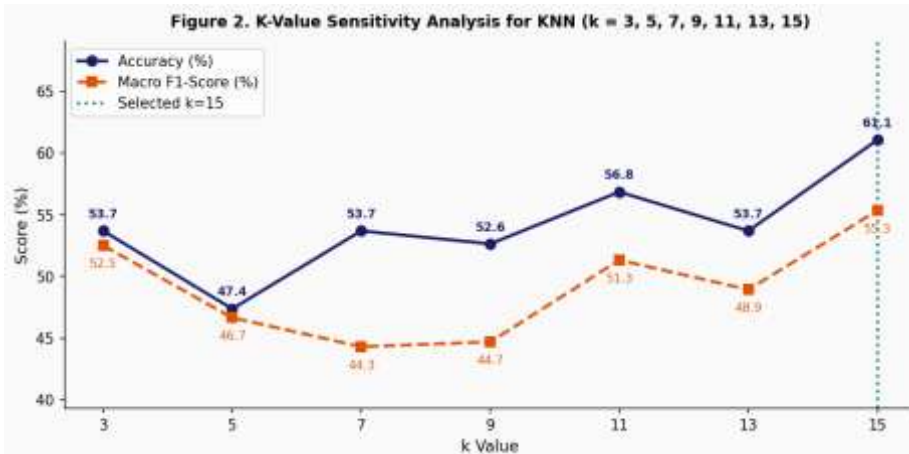


Fig. 2 K-Value Sensitivity Analysis for KNN ($k = 3, 5, 7, 9, 11, 13, 15$)

Hold-Out Testing Results

Table 3 presents the complete performance comparison across all four models on the hold-out test set ($n=95$).

Table 3. Complete Performance Comparison (Hold-Out Testing, $n=95$)

Model	Acc	Prec	Rec	F1	Bal.Acc	Macro F1
Majority Class Baseline	56.84%	56.84%	100.00%	72.48%	50.00%	36.24%
Naive Bayes (BernoulliNB)	56.84%	61.02%	66.67%	63.72%	55.28%	55.24%
KNN ($k=15$)	61.05%	61.33%	85.19%	71.32%	57.23%	55.33%
Decision Tree (depth=5)*	60.00%	62.50%	74.07%	67.80%	57.77%	57.51%

Highest Balanced Accuracy and Macro F1-Score

KNN ($k=15$) achieved the highest hold-out accuracy (61.05%) and recall (85.19%), but at the cost of high false positives ($FP=29$). Decision Tree achieved the highest balanced accuracy (57.77%) and macro F1-score (57.51%). Naive Bayes achieved identical hold-out accuracy to the baseline (56.84%) on this partition but substantially outperformed it on macro F1 (+19 pp). All three models substantially outperformed the majority class baseline on balanced accuracy (+5–8 pp) and macro F1-score (+19–21 pp), confirming meaningful learning beyond naive prediction.

Confusion Matrix

Table 4 presents the confusion matrix summary for all models on the hold-out test set. Fig. 3 provides a visual representation of the confusion matrices.

Table 4. Confusion Matrix Summary (Hold-Out Test Set, $n=95$)

Model	TN	FP	FN	TP
Majority Class Baseline	0	41	0	54
Naive Bayes (BernoulliNB)	18	23	18	36
KNN ($k=15$)	12	29	8	46
Decision Tree (depth=5)	17	24	14	40

*name of corresponding author



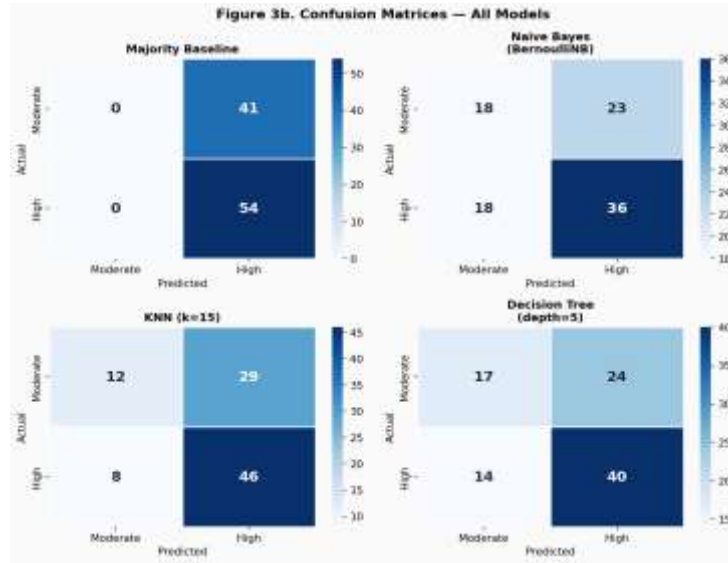


Fig. 3 Confusion Matrices for All Models (Test Set, n=95)

Performance Comparison

Fig. 4 illustrates the comparison of Accuracy, Balanced Accuracy, and Macro F1-Score across all four models. The chart clearly shows that while KNN leads on raw accuracy, Decision Tree leads on the imbalance-aware metrics (Balanced Accuracy and Macro F1), and all three trained models substantially outperform the majority class baseline.

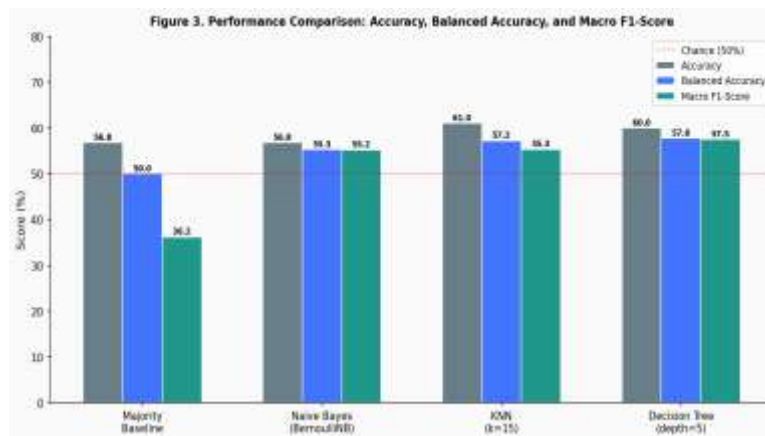


Fig. 4 Performance Comparison: Accuracy, Balanced Accuracy, and Macro F1-Score

Cross-Validation Results

Table 5 presents 10-fold cross-validation results, which provide a more robust estimate of generalization performance than a single hold-out partition (Brodersen et al., 2010). Fig. 5 shows the per-fold accuracy across all models.

Table 5. Cross-Validation Summary (10-Fold, n=471)

Model	CV Mean Accuracy	Std. Deviation
Majority Class Baseline	56.90%	0.87%
Naive Bayes (BernoulliNB)**	60.07%	6.02%
KNN (k=15)	54.78%	4.88%
Decision Tree (depth=5)	55.83%	5.60%

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Highest cross-validation mean accuracy

Naive Bayes demonstrated the highest CV mean accuracy (60.07%), suggesting superior generalization stability. KNN (k=15), despite achieving the highest hold-out accuracy, showed a substantial gap between hold-out (61.05%) and CV performance (54.78%) a 6.27 pp difference indicating sensitivity to the specific train-test partition. Decision Tree similarly showed a 4.17 pp gap (60.00% hold-out vs. 55.83% CV). Naive Bayes showed the smallest gap (3.23 pp: CV mean exceeds hold-out accuracy), confirming distributional consistency across folds.

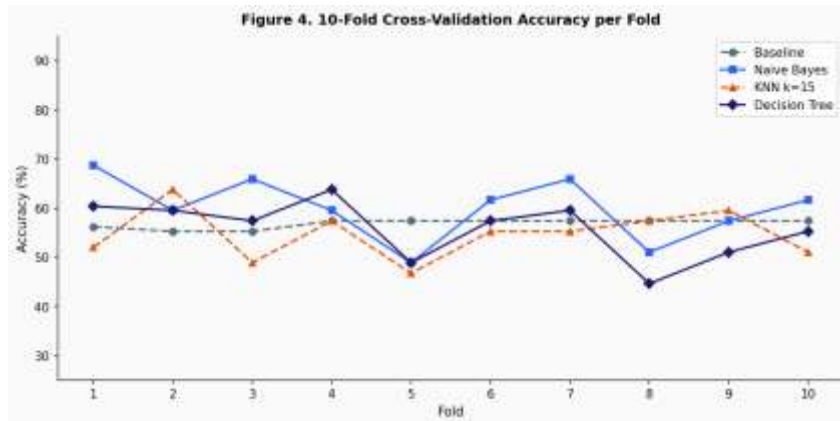


Fig. 5 10-Fold Cross-Validation Accuracy per Fold

Feature Importance Analysis

Fig. 6 presents the top 10 Decision Tree feature importance scores. IPS prior major contributed 15.6% of total importance, followed by Angkatan 2020 cohort (14.4%), IPA prior major (10.8%), and Sistem Informasi program (10.4%). Among the three feature categories, Jurusan (prior major) features collectively accounted for the largest share of importance. The presence of Angkatan 2020 as the second-highest contributor confirms that cohort membership adds meaningful temporal discriminative information independent of study program affiliation.

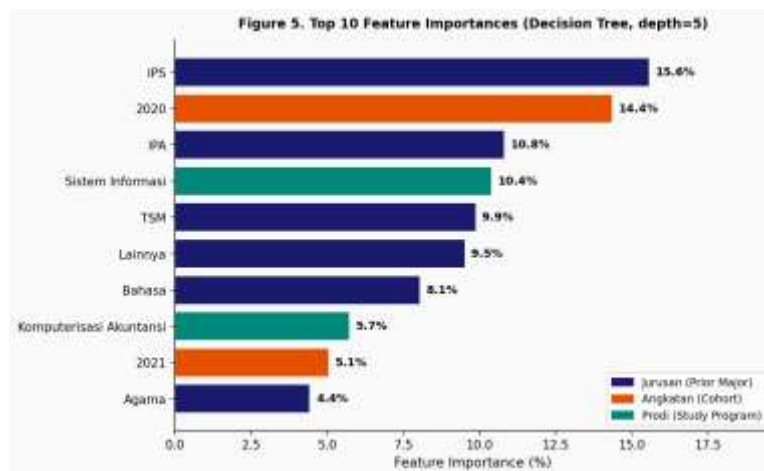


Fig. 6 Top 10 Feature Importances (Decision Tree, depth=5)

Per-Study-Program Analysis

Table 6 presents classification accuracy broken down by study program for programs with sufficient test set sample sizes ($n \geq 10$).

Table 6. Per-Study-Program Classification Accuracy

Study Program	n	High (%)	Naive Bayes	KNN (k=15)	Decision Tree
Sistem Informasi	55	58.2%	60.0%	61.8%	61.8%
Teknik Informatika	36	52.8%	52.8%	61.1%	58.3%

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

DISCUSSIONS

The results present a nuanced picture that extends beyond a simple algorithmic ranking, connecting to three analytically important themes: (1) the divergence between hold-out and cross-validation performance, (2) the structural constraints imposed by sparse binary feature spaces, and (3) the comparative value of interpretable versus predictive classification in institutional contexts.

Hold-Out vs. Cross-Validation Divergence and Its Implications

The divergence between hold-out and CV accuracy is most pronounced for KNN ($k=15$): 61.05% hold-out vs. 54.78% CV a 6.27 pp gap indicating that KNN's highest hold-out accuracy reflects sensitivity to the specific 80:20 partition rather than robust distributional learning. Decision Tree shows a 4.17 pp gap (60.00% vs. 55.83%), while Naive Bayes shows the smallest gap (3.23 pp). Notably, Naive Bayes CV mean (60.07%) actually exceeds its hold-out accuracy an indicator that NB's generalization improves with more training data, making it the most stable algorithm when evaluated across varying data partitions.

This distinction has direct institutional implications. Compared to Sarker et al. (2024), who reported NB accuracy of 70–85% on behaviorally rich datasets, the present study's 60.07% CV mean is substantially lower. This gap is directly attributable to the categorical-records-only condition rather than algorithmic limitations confirming Angeioplastis et al.'s (2025) finding that feature quality is more influential than algorithm selection. For institutions deploying a classifier updated periodically as new student cohorts arrive, CV accuracy is the more decision-relevant metric. Under this criterion, Naive Bayes with the highest CV mean (60.07%) and smallest hold-out/CV gap is the recommended algorithm for the categorical-records-only condition.

Structural Constraints of Sparse Binary Feature Spaces

The per-program results reveal an interesting exception to the overall pattern: Naive Bayes achieves only 52.8% accuracy on Teknik Informatika students the worst among the three classifiers for this program. Students in Teknik Informatika disproportionately originate from technical prior major backgrounds (RPL, TKJ), resulting in encoded vectors that share many common binary entries. Under this condition, the class-conditional probability estimates for Naive Bayes may be poorly calibrated due to the dominance of shared feature patterns, while KNN (61.1%) and Decision Tree (58.3%) handle program-specific concentrations more gracefully through distance-based and rule-based partitioning respectively (Cover & Hart, 1967; Angeioplastis et al., 2025).

This finding offers a theoretically grounded caution: the superiority of Naive Bayes for generalization (CV stability) does not translate uniformly across program subgroups. Institutions with strong program-specific feature concentrations should consider stratified retraining or program-specific model deployment.

The Interpretability Dividend

The feature importance analysis from Decision Tree produces findings that are institutionally actionable independent of the accuracy level achieved. The identification of IPS prior major (15.6%) and Angkatan 2020 cohort (14.4%) as the most discriminative features implies that students from social science backgrounds and those enrolled in 2020 have systematically different achievement distributions. Academic advisors can use this information to design targeted orientation or mentoring programs for specific Jurusan and cohort groups (Caday et al., 2025; Angeioplastis et al., 2025).

A critical caveat applies: the moderate accuracy ceiling (55–61% depending on metric and model) means feature importance scores should be treated as hypothesis-generating rather than causally definitive. The relatively even distribution across features with no single attribute exceeding 16% importance also confirms that no single background characteristic dominates achievement classification under the categorical-only condition.

Limitations and Future Directions

Several limitations constrain the generalizability of findings. The feature set was deliberately restricted to categorical background variables, excluding behavioral predictors consistently identified as strong in the literature (Nayak, 2023; Malik et al., 2025). The binary IPK threshold at 3.50 reduces output granularity. The single-institution dataset ($n=471$) limits external validity, and the benchmark requires replication across multiple institutions. Future research should incorporate behavioral and course-level features to quantify marginal predictive value, explore multiclass achievement frameworks, and evaluate ensemble methods such as Random Forest and Gradient Boosting, which demonstrate substantially stronger performance in feature-rich contexts (Nayak, 2023).

CONCLUSION

This study compared Naive Bayes (BernoulliNB), K-Nearest Neighbor (KNN), and Decision Tree against a majority class baseline for classifying student academic achievement under the categorical-records-only condition three background categorical attributes transformed into 18 binary features from 471 student records.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Three principal findings emerged. First, all three classifiers demonstrated meaningful learning beyond the majority class baseline, with +5–8 pp improvements on balanced accuracy and +19–21 pp on macro F1-score, confirming that institutional background attributes carry genuine class-discriminative signal even under feature constraints. Second, despite achieving the highest hold-out accuracy (61.05%), KNN showed the largest hold-out/CV gap (6.27 pp), making Naive Bayes with the highest CV mean (60.07%) and smallest divergence the most generalizable and deployment-appropriate algorithm for institutions retraining models on evolving student cohorts. Third, feature importance analysis identified IPS prior major (15.6%) and Angkatan 2020 cohort (14.4%) as the most discriminative features, providing actionable institutional intelligence for targeted academic interventions.

These findings establish the first documented performance benchmark for the categorical-records-only condition in Indonesian private higher education, provide evidence-based algorithm selection guidance for data-constrained institutions, and define a baseline against which future feature-enriched models can measure the marginal value of additional data collection investment. Future research is recommended to combine more advanced feature engineering with ensemble methods and contextual models to achieve more comprehensive performance improvements.

REFERENCES

- Agyemang, E. F. (2024). Predicting students' academic performance via machine learning classification: An empirical review and practical application. *International Journal of Advanced Computer Science and Applications*, 15(3), 512–521. <https://doi.org/10.14569/IJACSA.2024.0150358>
- Ahmed, E. (2024). Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, 2024(1), Article 4067721. <https://doi.org/10.1155/2024/4067721>
- Alalawi, K., Athauda, R., Chiong, R., & Renner, I. (2025). Evaluating the student performance prediction and action framework through a learning analytics intervention study. *Education and Information Technologies*, 30, 3891–3917. <https://doi.org/10.1007/s10639-024-12923-5>
- Angeioplastis, A., Aliprantis, J., Konstantakis, M., & Tsimpiris, A. (2025). Predicting student performance and enhancing learning outcomes: A data-driven approach using educational data mining techniques. *Computers*, 14(3), Article 83. <https://doi.org/10.3390/computers14030083>
- Boujmiraz, S., Darhmaoui, H., & Drissi El Maliani, A. (2026). Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches. *Computers and Education: Artificial Intelligence*, 10, Article 100548. <https://doi.org/10.1016/j.caeai.2026.100548>
- Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)* (pp. 3121–3124). IEEE. <https://doi.org/10.1109/ICPR.2010.764>
- Caday, M. B., Dalaorao, G. A., & Abalorio, C. C. (2025). Academic performance analysis of diploma students using predictive analytics. *Procedia Computer Science*, 257, 913–920. <https://doi.org/10.1016/j.procs.2025.03.117>
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*, 47(8), 5260–5267. <https://doi.org/10.1016/j.matpr.2021.05.646>
- Jannah, D. M., Susanto, A., Nugroho, H., & Prabowo, W. (2025). KNN vs Naive Bayes: An innovative comparison for prediction tasks in educational data. *Journal of Digital Technology and Computer Science*, 4(1), 45–56.
- Malik, S., Patro, S. G. K., Mahanty, C., Hegde, R., Naveed, Q. N., Lasisi, A., Buradi, A., Emma, A. F., & Kraiem, N. (2025). Advancing educational data mining for enhanced student performance prediction: A fusion of feature selection algorithms and classification techniques with dynamic feature ensemble evolution. *Scientific Reports*, 15, Article 8738. <https://doi.org/10.1038/s41598-025-92324-x>
- Nayak, P., Vaheed, S., Gupta, S., & Mohan, N. (2023). Predicting students' academic performance by mining the educational data through machine learning-based classification model. *Education and Information Technologies*, 29, 5685–5706. <https://doi.org/10.1007/s10639-023-11706-8>
- Powers, D. M. W. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *ArXiv preprint*. <https://arxiv.org/abs/2010.16061>
- Sarker, S., Paul, M. K., Thasin, S. T. H., & Hasan, M. A. M. (2024). Analyzing students' academic performance using educational data mining. *Computers and Education: Artificial Intelligence*, 7, Article 100263. <https://doi.org/10.1016/j.caeai.2024.100263>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

scikit-learn developers. (2024). *scikit-learn: Machine learning in Python* (Version 1.4). <https://scikit-learn.org/stable/>.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.