

Academic Risk Analysis of University Students Using Random Forest and SHAP

Nouval Khairi^{1)*}, Reza Ade Putra²⁾

^{1,2)} Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Indonesia

¹⁾nouvalkhairi123@gmail.com, ²⁾rezaadeputra@uinsu.ac.id

ABSTRACT

Background: Academic risk classification can help universities identify students who require closer academic monitoring, but imbalanced classes and limited model transparency can reduce practical usefulness. Objective: This study develops a Random Forest classifier for academic-risk categories and explains model decisions using SHAP. Methods: The study used academic data from 484 Computer Science students at Universitas Islam Negeri Sumatera Utara for the 2025/2026 odd-to-even academic period. The target was formed using rule-based indicators, while the final predictors were Jumlah_MK (number of courses), Praktikum (practicum courses), and MK_Berat (heavy courses). Data were split stratified into 387 training and 97 testing observations. SMOTE was applied only to training folds during stratified five-fold cross-validation and again to the complete training set for the final model. Results: Cross-validation produced mean Accuracy 91.72%, Precision 46.05%, Recall 81.33%, F1-Score 58.21%, and ROC-AUC 90.46%. On the original testing set, the model achieved Accuracy 97.94%, Precision 85.71%, Recall 85.71%, F1-Score 85.71%, and ROC-AUC 98.65%, with TN=89, FP=1, FN=1, and TP=6. Feature Importance and Mean Absolute SHAP both ranked Jumlah_MK as the most influential predictor. Conclusion: Random Forest with training-only SMOTE provided strong testing performance and improved minority-class detection, while SHAP added an interpretable layer. The results should be interpreted cautiously because the high-risk class was small and the study used a single academic period.

Keywords: academic risk; explainable artificial intelligence; Random Forest; SHAP; SMOTE; student analytics

INTRODUCTION

Universities increasingly use academic data to monitor student progress and to support timely academic intervention. Indicators such as grade performance, study load, course completion, and other registration information can be transformed into analytical features for classifying students according to academic conditions. Prior evidence shows that academic achievement and dropout-related outcomes are associated with multiple interacting factors, while machine learning can model non-linear patterns from student records (Kocsis & Molnár, 2025; Nurmawati et al., 2023). In this context, classification of academic risk can complement conventional monitoring by organizing students into risk categories that can be examined by academic staff.

Random Forest is a suitable candidate for this problem because it combines multiple decision trees and can model complex feature interactions. Previous studies have reported strong performance of Random Forest for student academic classification, graduation, and dropout-related tasks (Zawiyah et al., 2024; Sulehu et al., 2025; Morán et al., 2025). However, performance metrics can be misleading when the target distribution is highly imbalanced. A model may obtain high accuracy while failing to detect a small minority class. Synthetic Minority Over-sampling Technique (SMOTE) is therefore relevant for increasing the representation of minority observations during model training. Importantly, oversampling should be restricted to the training portion so that validation and testing remain independent of synthetic observations.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

A second issue is interpretability. Although tree ensembles can perform well, a prediction alone does not explain why a particular student is classified into a risk category. Explainable Artificial Intelligence (XAI) addresses this limitation by providing explanations of model outputs. SHAP represents one widely used approach for estimating the contribution of individual predictors, both at the global level and for individual observations (Lundberg & Lee, 2017; Er, 2023; Antonini et al., 2024). Recent studies have applied explainable machine learning in student-related contexts, including academic-risk analysis and behavioral classification (Ananda et al., 2024; Hidayatulloh et al., 2026; Pajung et al., 2026).

The research gap addressed here lies in the combination of a context-specific academic-risk definition, minority-class handling, and model interpretation while separating target-forming variables from model predictors. In the present dataset, IP, SKS_Diambil, and Rasio_SKS are used in forming the target label and are therefore excluded from the final feature matrix to reduce target-feature circularity. The model instead uses the study-load characteristics Jumlah_MK, Praktikum, and MK_Berat. The contribution of this study is a reproducible Random Forest pipeline with stratified five-fold cross-validation, fold-level SMOTE, final testing on untouched data, and SHAP-based global and local interpretation. The objective is to evaluate the classification capability of Random Forest and to identify which study-load features contribute most to its decisions.

LITERATURE REVIEW

Research on academic analytics shows that machine learning has been applied to dropout, graduation, academic performance, and learning behavior. Morán et al. (2025) used Random Forest to identify factors related to student dropout, reporting an accuracy of 79%, Precision 82%, Recall 73%, and AUC 0.86. Zawiyah et al. (2024) reported 98.2% accuracy for academic-performance classification using Random Forest. Other studies have compared tree-based models for graduation or dropout prediction and have highlighted the importance of robust evaluation across multiple metrics rather than relying on accuracy alone (Junaidi et al., 2024; Putra et al., 2025; Saputra, 2026).

Class imbalance is particularly relevant when the risk class represents a smaller proportion of students. Pratiwi and Utami (2025) applied Random Forest with SMOTE in student-performance prediction, while the present study applies SMOTE specifically within the training side of each cross-validation fold and keeps the validation and testing observations in their original distributions. This design is intended to reduce information leakage and to make the reported validation and testing metrics more representative of unseen data.

Interpretability provides an additional layer of scientific value. SHAP-based explanations have been applied to academic-risk, academic-performance, and related classification tasks (Ananda et al., 2024; Hidayatulloh et al., 2026; Tamsir et al., 2026). In this study, Mean Absolute SHAP is used for global ranking, while a local explanation is used to illustrate how the predictors jointly influence a selected high-risk observation. Unlike causal analysis, SHAP is treated as an explanation of model behavior.

METHOD

This study uses a quantitative experimental design based on structured academic records. The dataset contains 484 Computer Science students from cohorts 2022, 2023, and 2024 at Universitas Islam Negeri Sumatera Utara. The data were extracted from Kartu Rencana Studi (KRS) records and integrated into a tabular dataset. Information used to construct the target included IP, SKS_Diambil, and Rasio_SKS, with Tahap_Akhir used as supporting information to prevent students in the study-completion stage from being classified as high risk solely because their current course load was low. The final predictors were Jumlah_MK, Praktikum, and MK_Berat.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 1. Dataset Characteristics

Item	Value
Total observations	484
Risk Low	449 (92.77%)
Risk High	35 (7.23%)
Final predictors	Jumlah_MK, Praktikum, MK_Berat
Target construction variables	IP, SKS_Diambil, Rasio_SKS

The target was produced using a rules-based labeling procedure. A student was assigned to the high-risk class when at least one of the following conditions was met: $IP < 2.75$; $SKS_Diambil < 18$ while not in the final-study stage; or $Rasio_SKS < 0.75$ while not in the final-study stage. Students not meeting these conditions were assigned to the low-risk class. The resulting target distribution is highly imbalanced, with 449 low-risk and 35 high-risk observations.

A stratified 80:20 train-test split was performed with `random_state = 42`, producing 387 training observations and 97 testing observations. The training set contained 359 low-risk and 28 high-risk observations, while the testing set contained 90 low-risk and 7 high-risk observations. The testing set was isolated before cross-validation and was not used in SMOTE. Stratified five-fold cross-validation was then conducted on the training set. Within every fold, SMOTE was applied only to the training fold, leaving the validation fold unchanged. After cross-validation, SMOTE was applied to the full training set to construct the final training data, resulting in 359 low-risk and 359 high-risk observations.

Feature construction was derived from the KRS records rather than from individual course grades. `Jumlah_MK` represents the number of registered courses, `Praktikum` counts courses identified as practicum subjects, and `MK_Berat` counts courses classified as academically demanding according to the study's predefined heavy-course list. `Tahap_Akhir` was used only as a contextual variable for target formation. Students taking Seminar Proposal, Tugas Akhir 1, Tugas Akhir 2, and/or Skripsi were treated as being in the study-completion stage. This rule prevents a low current study load caused by completion of most coursework from being interpreted automatically as academic risk. The final modeling matrix therefore contains only three predictors: `Jumlah_MK`, `Praktikum`, and `MK_Berat`. This design makes the model interpretation more closely aligned with the information that is available as independent study-load evidence rather than with the variables that define the label itself.

Table 2. Data Partition and Balancing

Stage	Low Risk	High Risk	Total
Training before SMOTE	359	28	387
Training after SMOTE	359	359	718
Testing	90	7	97

The Random Forest model used 100 trees, Gini impurity as the split criterion, `max_depth = None`, `min_samples_split = 2`, `min_samples_leaf = 1`, `max_features = sqrt`, and `random_state = 42`. No hyperparameter tuning was performed. The model was evaluated using Accuracy, Precision, Recall, F1-Score, ROC-AUC, and a confusion matrix. Feature Importance was extracted from the fitted Random Forest, while SHAP was used for global and local interpretation.

For model evaluation, Accuracy measures the proportion of all correctly classified observations, whereas Precision and Recall focus on the high-risk class. Recall is especially important in this study because a false negative corresponds to a student who is actually labeled high risk but is classified as low risk. F1-Score summarizes the trade-off between Precision and Recall, while ROC-AUC evaluates the ranking ability of the classifier across probability thresholds. The confusion matrix is used to retain the raw

counts of true and false predictions. These metrics are calculated on validation folds that remain in their original class distribution and on the untouched testing set. Consequently, synthetic SMOTE observations affect only model fitting and do not enter the reported final evaluation. The same modeling configuration is retained across folds and the final training stage so that differences in performance primarily reflect the data partition and balancing procedure rather than changes in model specification.

The analytical design separates target construction from predictor modeling to avoid target-feature circularity. The academic records first undergo structural checking, including verification of student identity fields, detection of duplicate records, and inspection of missing values. Repeated student names are not treated as duplicate observations when the corresponding student identifiers differ. Missing values in the supporting field MataKuliah_Tahap_Akhir are retained because that field is not included in the final feature matrix. The purpose of this separation is important because IP, SKS_Diambil, and Rasio_SKS participate directly in the rules used to define Label_Final; reintroducing those same variables as predictors would allow the model to reproduce part of the labeling rule rather than learn an independent pattern from study-load characteristics.

RESULT

The class distribution confirms that the low-risk category dominates the dataset. Figure 1 illustrates the imbalance, which is the main reason that accuracy alone is insufficient for model evaluation. The minority class contains only 35 observations before the train-test split, and only seven high-risk observations appear in the final testing set.

Before model validation, the three final predictors were summarized to describe their empirical range and central tendency. The descriptive statistics show that the observed study-load variables have meaningful variation across students, providing the basis for classification without directly using the target-forming variables.

Table 3. Descriptive Statistics of Final Predictors

Predictor	Mean	Median	Std. Dev.	Minimum	Maximum
Jumlah_MK	9.6364	12.0000	4.5661	0	15
Praktikum	2.6302	3.0000	1.7331	0	7
MK_Berat	1.603	1,00	1.236	0	3

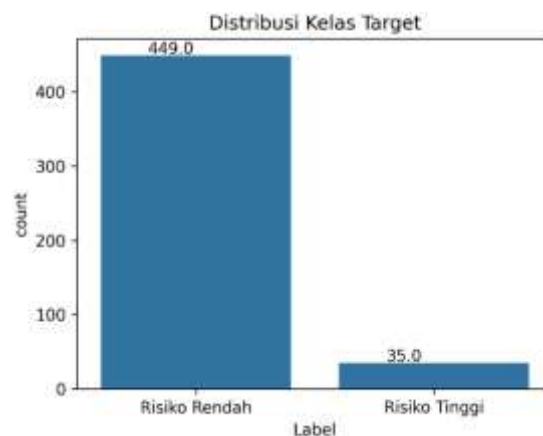


Fig 1. Distribution of Academic-Risk Classes

The stratified five-fold cross-validation results are summarized in Table 4. Accuracy ranged from 89.61%

* Corresponding author



to 93.59%, indicating relatively low fold-to-fold variation. Recall ranged from 60.00% to 100.00%, with a mean of 81.33% and a standard deviation of 18.50%. The large recall variation is consistent with the small number of high-risk observations available within the training folds.

Table 4. Stratified Five-Fold Cross-Validation Results

Fold	Accuracy	Precision	Recall	F1-Score	ROC-AUC
1	93.59%	54.55%	100.00%	70.59%	97.57%
2	92.31%	50.00%	66.67%	57.14%	80.32%
3	90.91%	40.00%	80.00%	53.33%	88.75%
4	92.21%	42.86%	60.00%	50.00%	86.81%
5	89.61%	42.86%	100.00%	60.00%	98.83%
Mean	91.72%	46.05%	81.33%	58.21%	90.46%
Std. Dev.	1.52%	6.02%	18.50%	7.89%	7.74%

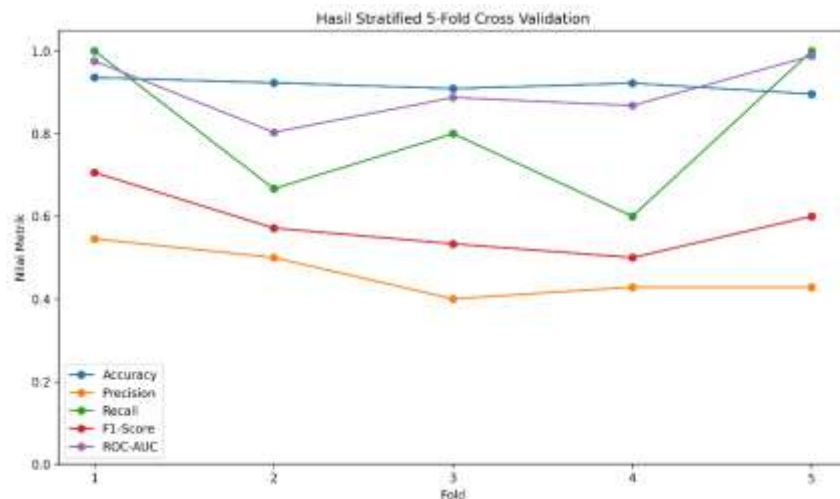


Fig 2. Stratified Five-Fold Cross-Validation Performance

The fold-level results also show why the mean values should be interpreted jointly with dispersion. Validation Accuracy varied from 89.61% to 93.59%, while Recall varied from 60.00% to 100.00%. The training data contained only 28 high-risk observations before SMOTE, so each validation fold contained a small number of positive cases. A change of one or two predictions can therefore cause a noticeable change in Recall and F1-Score. The mean Accuracy of 91.72% has a comparatively small standard deviation of 1.52%, whereas Recall has a standard deviation of 18.50%. This difference indicates that overall correctness is more stable than minority-class detection. The result is consistent with the objective of reporting class-sensitive metrics in an imbalanced setting rather than treating high Accuracy as sufficient evidence of a reliable risk classifier.

After validation, the final Random Forest was trained using the SMOTE-balanced training data and evaluated on the untouched 97-observation testing set. The confusion matrix produced TN = 89, FP = 1, FN = 1, and TP = 6. Thus, 95 of 97 observations were classified correctly, while one low-risk observation was predicted as high risk and one high-risk observation was predicted as low risk.

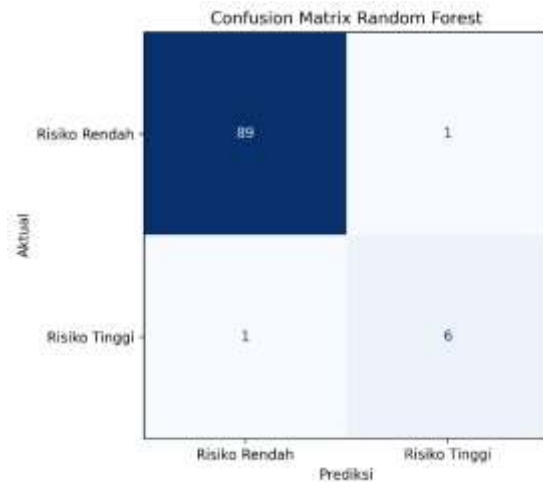


Fig 3. Confusion Matrix of the Final Random Forest

Table 5. Final Testing Performance

Metric	Value
Accuracy	97.94%
Precision	85.71%
Recall	85.71%
F1-Score	85.71%
ROC-AUC	98.65%

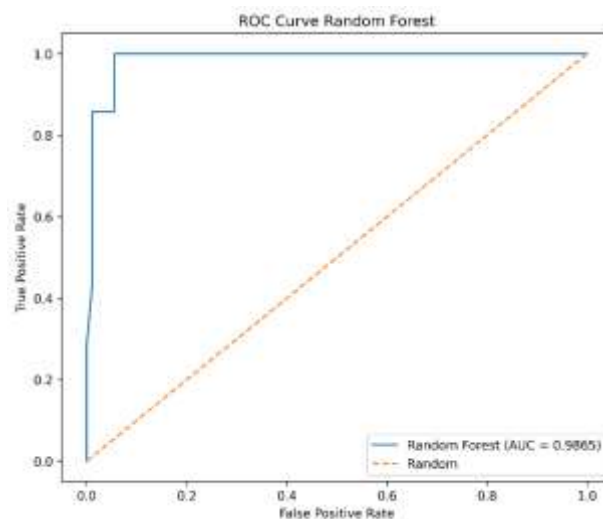


Fig 4. ROC Curve of the Final Random Forest

The testing performance indicates that the model detected six of the seven high-risk students. The testing Recall of 85.71% is therefore driven by a very small positive class, and a single additional error would materially change the metric. The ROC-AUC of 98.65% indicates strong ranking discrimination, but it should not be interpreted as the percentage of correctly classified observations.

Because the Random Forest uses the Gini criterion, the class proportions before balancing also provide a useful description of the impurity encountered at the start of training. The original training distribution of

* Corresponding author



359 low-risk and 28 high-risk observations gives a Gini impurity of approximately 0.1342. This value describes class impurity at a node and is not a model-performance metric. It should therefore be distinguished from Accuracy, Precision, Recall, F1-Score, and ROC-AUC, which are computed from model predictions on validation or testing observations.

The effect of SMOTE was assessed by comparing otherwise equivalent Random Forest configurations. Without SMOTE, the testing Accuracy was 95.88%, Precision 100.00%, Recall 42.86%, F1-Score 60.00%, and ROC-AUC 98.81%. With SMOTE, Accuracy increased to 97.94%, Recall to 85.71%, and F1-Score to 85.71%, while Precision declined to 85.71% and ROC-AUC changed slightly to 98.65%. The main practical benefit was the stronger detection of the minority high-risk class.

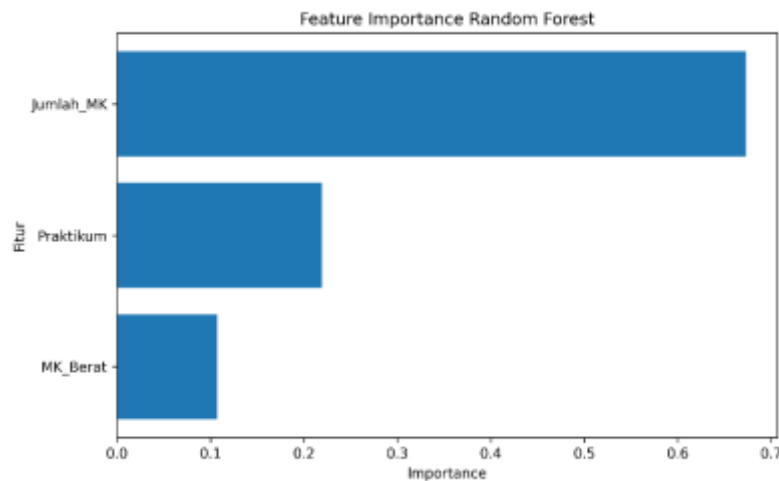


Fig 5. Random Forest Feature Importance

Feature Importance ranked Jumlah_MK first (0.673296), followed by Praktikum (0.219303) and MK_Berat (0.107401). SHAP produced the same global ranking. Mean Absolute SHAP was 0.337941 for Jumlah_MK, 0.124441 for Praktikum, and 0.094799 for MK_Berat. The consistency between the two global importance approaches strengthens the interpretation that Jumlah_MK is the dominant predictor in the fitted model.

At the local explanation level, one correctly identified high-risk observation had Jumlah_MK = 10, Praktikum = 4, and MK_Berat = 0. For the analyzed class, the SHAP base value was 0.5061. Jumlah_MK contributed +0.4352, while Praktikum and MK_Berat contributed -0.2141 and -0.1043, respectively, producing an output of approximately 0.6228 in the explainer's output space. This example illustrates the additive nature of SHAP: the final explanation is obtained by combining a baseline with feature-level contributions. The result does not mean that a particular value of a predictor causes academic risk; it only describes how that value shifted the fitted model output for the selected observation.

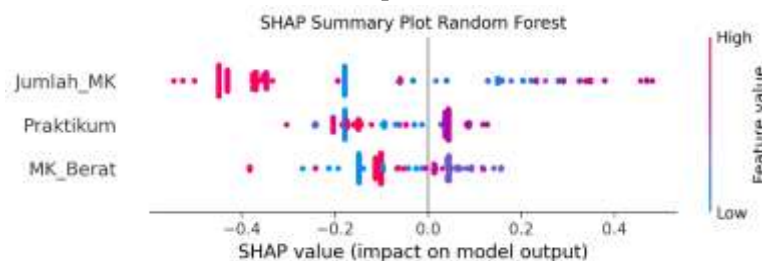


Fig 6. SHAP Summary Plot for the Final Model

DISCUSSIONS

The Random Forest results show a clear difference between validation and final testing. The mean cross-validation Accuracy was 91.72%, while the testing Accuracy reached 97.94%. More importantly for the minority class, Recall rose from 81.33% in cross-validation to 85.71% in testing. This difference should not be treated as direct evidence of overfitting because the testing set is independent and contains only seven high-risk observations. With such a small minority class, a change of one or two predictions can cause a large change in Recall, Precision, and F1-Score. The relatively large cross-validation standard deviation for Recall (18.50%) confirms that minority-class detection is not fully stable across folds.

The SMOTE comparison provides a more direct interpretation of the balancing procedure. The principal change is not a universal increase in every metric; rather, it is a substantial increase in minority-class sensitivity. Recall improved by 42.86 percentage points and F1-Score by 25.71 percentage points, while Precision decreased by 14.29 percentage points. This trade-off is coherent with the purpose of oversampling in an imbalanced setting: the classifier becomes more willing to predict the minority class, thereby reducing false negatives at the cost of some additional false positives. In the context of academic-risk screening, reducing false negatives can be valuable because a high-risk student classified as low risk may not receive further attention.

The results are broadly consistent with prior studies that demonstrate the usefulness of Random Forest for student-related classification. Zawiyah et al. (2024) reported high accuracy for academic-performance classification, while Morán et al. (2025) reported lower but more balanced performance across accuracy, Precision, Recall, and AUC in dropout-oriented prediction. The present study differs in target definition and feature design. It classifies a rule-based academic-risk category for a single academic period rather than predicting a future dropout or graduation outcome. It also excludes the variables used directly in target construction, which reduces the risk of target-feature circularity but leaves a narrower feature space.

The SHAP analysis adds interpretability that is not available from accuracy metrics or tree-level importance alone. Jumlah_MK consistently had the largest global contribution, while the local explanation for a selected high-risk observation showed a positive SHAP contribution from Jumlah_MK and negative contributions from Praktikum and MK_Berat. These values describe how the fitted model combines features for the selected observation; they do not establish that increasing or decreasing a feature causes academic risk. This distinction is important when translating model explanations into academic intervention policies.

From an academic-monitoring perspective, the most useful aspect of the model is not its Accuracy alone but the combination of minority-class sensitivity and an explicit explanation layer. The SMOTE experiment demonstrates this trade-off clearly: Recall increased from 42.86% to 85.71%, while Precision decreased from 100.00% to 85.71%. The change indicates that the balanced training procedure reduced the tendency to ignore the minority class. In a screening context, this can be preferable to a model that achieves high overall Accuracy by favoring the majority class, because false negatives may leave potentially high-risk students without additional attention.

Several limitations should be considered. First, the high-risk class is small, resulting in substantial sensitivity of minority-class metrics to individual observations. Second, the dataset represents one academic period, so the study does not establish future predictive validity for subsequent semesters, graduation, or dropout. Third, no hyperparameter tuning was performed, meaning the reported configuration is an explicit experimental setting rather than an optimized configuration. Future studies should use longitudinal outcomes, a larger minority class, additional predictors that are independent of target construction, controlled hyperparameter tuning, and fair comparison with alternative classifiers.

The findings also suggest several considerations for practical use. First, a classification model intended to support academic monitoring should be positioned as a screening instrument rather than an automatic decision-maker. The predicted high-risk category can indicate which observations deserve further review,

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

but academic staff should still consider contextual information that is not represented in the current dataset. Second, the explanation layer can support a more transparent workflow because staff can inspect which predictors contributed most strongly to a model output. In this study, the agreement between Random Forest Feature Importance and Mean Absolute SHAP provides a consistent global view of the fitted model, while the local SHAP example illustrates how the same predictors can contribute in different directions for an individual observation. This combination is more informative than presenting a single classification label without explanation.

Reproducibility is another important consideration. The study explicitly fixes the train-test split with `random_state = 42` and applies the same Random Forest configuration throughout the evaluation. The separation of the testing set before cross-validation, followed by fold-specific SMOTE, means that synthetic observations are confined to model training and cannot directly enter validation or final testing. This is a key part of the experimental design because the reported performance is intended to reflect classification on observations that were not used to generate synthetic training examples. The same principle should be preserved when extending the study with additional algorithms or tuning procedures. Any preprocessing, resampling, feature selection, or parameter search that can learn from the data should remain inside the training side of the evaluation pipeline.

The interpretation of the dominant predictor also requires contextual caution. `Jumlah_MK` has the highest Feature Importance and Mean Absolute SHAP in the fitted model, but the result should not be translated into a direct substantive claim that course count independently causes academic risk. The predictor may capture several underlying characteristics of a student's registration pattern, including study stage and course-load structure. Because the current target is constructed from academic indicators that are conceptually related to study progress, statistical association and model contribution should be kept separate from causal inference. A future longitudinal study could test whether the predictors available in one semester provide useful information about an objectively observed outcome in a later semester. Such a design would also make it possible to estimate prospective recall, precision, and calibration under conditions closer to an operational early-warning system.

Finally, the results should be interpreted at the scale of the available evidence. The testing set contains 97 students, but only seven belong to the high-risk class. The observed Recall of 85.71% corresponds to six detected high-risk cases, so the metric is strongly influenced by a very small number of observations. The same issue is visible in cross-validation, where Recall has a standard deviation of 18.50%. Therefore, the contribution of this study is best understood as a methodological demonstration of an interpretable and leakage-conscious classification pipeline on a specific academic dataset, rather than as evidence that the reported performance will generalize unchanged to other cohorts. Broader validation across semesters, cohorts, and institutions would be necessary before operational adoption.

The study also has methodological implications for educational data mining. Because the target is rules-based and the data represent one academic period, the model should be viewed as a classification aid rather than a validated early-warning predictor of future outcomes. The separation between target-forming variables and model predictors is a deliberate design decision that reduces circularity, but it also limits the information available to the classifier. Future work can preserve the same leakage-avoidance principle while adding longitudinal predictors and observed future outcomes, allowing performance to be assessed prospectively.

CONCLUSION

Random Forest successfully classified the academic-risk categories of 484 Computer Science students using three predictors: `Jumlah_MK`, `Praktikum`, and `MK_Berat`. With SMOTE restricted to training data and applied within each cross-validation training fold, the model achieved mean cross-validation Accuracy 91.72%, Precision 46.05%, Recall 81.33%, F1-Score 58.21%, and ROC-AUC 90.46%. On the untouched

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

testing set, it achieved Accuracy 97.94%, Precision 85.71%, Recall 85.71%, F1-Score 85.71%, and ROC-AUC 98.65%, with TN=89, FP=1, FN=1, and TP=6. SMOTE primarily improved high-risk detection, increasing Recall and F1-Score. Feature Importance and SHAP agreed that Jumlah_MK was the most influential predictor. The contribution of the study is a transparent classification workflow that combines imbalance handling and model explanation, while the results remain bounded by the small minority class and single-period design.

REFERENCES

- Ananda, L. F., Dewi, M. R., & Habibi, M. R. (2024). Explainable machine learning dalam analisis risiko akademis mahasiswa Fakultas Vokasi Institut Teknologi Sepuluh Nopember. *IJAI (Indonesian Journal of Applied Informatics)*, 9(1), 113–123. <https://doi.org/10.20961/ijai.v9i1.94691>
- Antonini, A. S., Tanzola, J., Asiain, L., Ferracutti, G. R., Castro, S. M., Bjerg, E. A., & Ganuza, M. L. (2024). Machine learning model interpretability using SHAP values: Application to igneous rock classification task. *Applied Computing and Geosciences*, 23, 100178. <https://doi.org/10.1016/j.acags.2024.100178>
- Er, E. (2023). An explainable machine learning approach to predicting and understanding dropouts in MOOCs. *Kastamonu Eğitim Dergisi*, 31(1), 143–154. <https://doi.org/10.24106/kefdergi.1246458>
- Firman, M. A., Djamalilleil, S. A. F., Zega, W., Efrizoni, L., & Rahmadden. (2025). Model klasifikasi IPK mahasiswa menggunakan algoritma Decision Tree dan Random Forest berbasis feature engineering. *TechnoCom*, 24(2). <https://doi.org/10.62411/tc.v24i2.12384>
- Fitriani, S., Budiman, E., Fadli, M., Surono, M., & Sulistiani, H. (2025). Optimalisasi metode Random Forest menggunakan particle swarm optimization dalam prediksi prestasi mahasiswa. *Prosiding Seminar Nasional Teknologi Komputer dan Sains (SAINTEKS)*, 3(1), 406–415.
- Gori, T., & Hestiningtyas, A. (2024). Optimasi pemilihan fitur untuk prediksi penyakit jantung menggunakan algoritma genetika dan Random Forest. *Jurnal Ilmu Komputer Indonesia*, 13(5). <https://doi.org/10.33022/ijcs.v13i5.4214>
- Hapsari, L. N., Fannani, I., Rahmawati, Y., & Muhariya, A. (2026). Explainable machine learning untuk prediksi risiko penyakit jantung menggunakan Random Forest dan analisis SHAP. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, 10(1), 190–199. <https://doi.org/10.33395/remik.v10i1.15766>
- Helmud, E., Fitriyani, F., & Romadiana, P. (2024). Classification comparison performance of supervised machine learning Random Forest and Decision Tree algorithms using confusion matrix. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 13(1), 92–97. <https://doi.org/10.32736/sisfokom.v13i1.1985>
- Hidayatulloh, W., Mahardika, F., & Junaedi, D. I. (2026). Explainable artificial intelligence-based model for student academic performance prediction. *Journal of Information System Exploration and Research*, 4(1), 31–40. <https://doi.org/10.52465/joiser.v4i1.624>
- Istiwana, A. P., Sani, R. R., & Pramudi, Y. T. C. (2026). Pendekatan explainable machine learning untuk analisis faktor drop out mahasiswa menggunakan XGBoost. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 11(1), 1074–1083. <https://doi.org/10.36341/rabit.v11i1.7218>
- Junaidi, S., Anggela, R. V., & Kariman, D. (2024). Klasifikasi metode data mining untuk prediksi kelulusan tepat waktu mahasiswa dengan algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Network (ANN). *JACOST: Journal of Applied Computer Science and Technology*, 5(1). <https://doi.org/10.52158/jacost.v5i1.489>
- Kocsis, Á., & Molnár, G. (2025). Factors influencing academic achievement and dropout rates in higher education. *Oxford Review of Education*, 51(3), 414–432. <https://doi.org/10.1080/03054985.2024.2316616>
- Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An

- example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, 96, 101845. <https://doi.org/10.1016/j.compenvurbsys.2022.101845>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.
- Morán, P. L., Torres, J. R., & Cruz, A. M. (2025). Identifying key factors of student dropout through Random Forest: A data-driven approach. *Applied Sciences*, 15(4), 1256. <https://doi.org/10.3390/app15041256>
- Nurmalitasari, Long, Z. A., & Noor, M. F. M. (2023). Factors influencing dropout students in higher education. *Education Research International*, 2023, Article 7704142. <https://doi.org/10.1155/2023/7704142>
- Pajung, K. K. M., Komansilan, T., Kawilarang, G. J. C., Thomas, M. J., Nonga, R. G. P., Sumeke, E. A., Kereh, G., & Kiwo, M. (2026). Klasifikasi perilaku belajar mahasiswa menggunakan Random Forest Classifier berbasis Behavior Questionnaire. *EduTIK: Jurnal Pendidikan Teknologi Informasi dan Komunikasi*, 6(2).
- Pratiwi, A. N., & Utami, E. (2025). Predicting students' academic performance in mathematics based on Big Five personality traits using Random Forest with Synthetic Minority Over-sampling Technique. *Sistemasi: Jurnal Sistem Informasi*, 14(2). <https://doi.org/10.32520/stmsi.v14i2.5102>
- Putra, A. P., Sari, N. P., & Wijaya, R. (2025). Student dropout prediction using Random Forest and XGBoost method. *Journal of Educational Data Mining and Analytics*, 7(2), 145–158.
- Saputra, P. S. (2026). Analisis prediktif dropout mahasiswa berdasarkan kinerja akademik semester awal menggunakan machine learning. *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, 7(1), 164–171. <https://doi.org/10.30998/ed38e865>
- Sulehu, M., Wisda, W., Wanita, F., & Markani. (2025). Optimasi prediksi kelulusan mahasiswa menggunakan Random Forest untuk meningkatkan tingkat retensi. *Jurnal Minfo Polgan*, 13(2), 2364–2374. <https://doi.org/10.33395/jmp.v13i2.14472>
- Tamsir, I. S., Putrada, A. G., & Wicaksono, R. L. (2026). SHAP dalam explainable AI untuk deteksi depresi mahasiswa menggunakan AdaBoost. *e-Proceeding of Engineering*, 13(3), 29661–29727.
- Zawiyah, S., Qodriyah, L., & Tamam, M. B. (2024). Klasifikasi prestasi akademik mahasiswa menggunakan metode Random Forest. *Jobit: Journal of Digital Business and Information Technology*, 1(2). <https://doi.org/10.23971/jobit.v1i2.317>
- Zulfa, E., Amir, H., Ginting, R., & Sudarno. (2024). Analisis korelasi kesehatan mental dan indeks prestasi mahasiswa Jurusan Administrasi Niaga Politeknik Negeri Jakarta dengan kombinasi metode XGBoost dan SHAP. *JAProf: Jurnal Administrasi Profesional*, 5(1), 26–37. <https://doi.org/10.32722/jap.v5i1.6923>