Random Forest vs XGBoost untuk Prediksi Risiko Drop Out Mahasiswa: Studi Eksploratif dan Implikasi bagi PTKI Indonesia

Hery Afriyadi[*1] Pol Metra[2], Hozana Aulia[3], Fatima Felawati[4]

[1234]Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi

E-mail: hery.afriyadi@uinjambi.ac.id[1], polmetra@uinjambi.ac.id[2], hozana.aulia@uinjambi.ac.id[3], fatimafelawati@uinjambi.ac.id[4]

Abstrak. Tingginya angka putus studi (dropout) merupakan tantangan dalam pengelolaan pendidikan tinggi sehingga diperlukan pendekatan prediktif untuk mengidentifikasi risiko secara dini. Penelitian ini membandingkan Random Forest dan XGBoost dalam memprediksi risiko dropout serta mengidentifikasi variabel dengan kontribusi prediktif terbesar menggunakan dataset *Predict Students' Dropout and Academic Success* dari UCI Machine Learning Repository, yang berasal dari Polytechnic Institute of Portalegre, Portugal. Setelah kelas *Enrolled* dikeluarkan, dataset terdiri atas 3.630 mahasiswa. Eksperimen menggunakan MinMaxScaler dan SMOTE, serta skema *nested cross-validation* dengan *inner loop* GridSearchCV 5-fold untuk *hyperparameter tuning* dan *outer loop* Repeated Stratified 5-fold × 10 pengulangan (50 evaluasi) untuk estimasi performa tak bias. XGBoost menunjukkan performa rata-rata sedikit lebih tinggi, dengan Accuracy 0,9085, Recall 0,8406, F1-Score 0,8779, dan ROC-AUC 0,9568. Uji Wilcoxon menunjukkan perbedaan signifikan pada Recall, F1-Score, dan ROC-AUC, tetapi tidak pada Accuracy maupun Precision. *Permutation feature importance* menunjukkan jumlah *curricular units* yang disetujui pada semester kedua sebagai fitur paling berpengaruh, diikuti status pembayaran uang kuliah. *Ablation study* menunjukkan penghapusan GDP, *inflation rate*, dan *unemployment rate* hanya memberikan perubahan performa kecil. Karena dataset berasal dari Portugal, hasil ini bersifat eksploratif dan belum dapat digeneralisasikan langsung pada Perguruan Tinggi Islam di Indonesia tanpa validasi eksternal.

Kata kunci: Dropout prediction; Random Forest; XGBoost; nested cross-validation; Permutation feature importance.

Abstract. *Student dropout is a challenge in higher education management, requiring predictive approaches for early risk identification. This study compares Random Forest and XGBoost for dropout prediction and identifies variables with the greatest predictive contribution using the Predict Students' Dropout and Academic Success dataset from the UCI Machine Learning Repository, collected at the Polytechnic Institute of Portalegre, Portugal. After excluding the Enrolled class, the dataset contained 3,630 students. The pipeline applied MinMaxScaler and SMOTE, together with a nested cross-validation scheme: an inner 5-fold GridSearchCV loop for hyperparameter tuning and an outer Repeated Stratified 5-fold × 10-repetition loop (50 evaluations) for unbiased performance estimation. XGBoost achieved slightly higher average performance, with Accuracy 0.9085, Recall 0.8406, F1-Score 0.8779, and ROC-AUC 0.9568. The Wilcoxon signed-rank test indicated significant differences in Recall, F1-Score, and ROC-AUC, but not in Accuracy or Precision. Permutation feature importance identified approved*

curricular units in the second semester as the most influential feature, followed by Tuition Fees Up to Date. Ablation analysis showed that removing GDP, inflation rate, and unemployment rate caused only minor performance changes. As the dataset was collected in Portugal, the findings are exploratory and cannot be directly generalized to Islamic Higher Education institutions in Indonesia without external validation.

Keywords: *Dropout prediction; Random Forest; XGBoost; nested cross-validation; Permutation feature Importance.*

1. Pendahuluan

Angka putus studi atau drop out (DO) mahasiswa merupakan salah satu indikator kualitas manajemen perguruan tinggi yang mendapat perhatian serius dari Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi (Ditjen Diktiristek). Data Pangkalan Data Pendidikan Tinggi (PDDikti) menunjukkan bahwa rata-rata angka DO mahasiswa di Indonesia berkisar antara 10–15% per tahun, dengan variasi yang cukup signifikan di antara jenis institusi [1]. Perguruan Tinggi Islam (PTI) yang dikelola di bawah koordinasi Kementerian Agama, khususnya yang berstatus Perguruan Tinggi Keagamaan Islam (PTKI), menghadapi tantangan tambahan terkait karakteristik mahasiswanya yang beragam, termasuk adanya mahasiswa yang juga berperan sebagai santri pondok pesantren[2].

Upaya pencegahan drop out secara konvensional umumnya bersifat reaktif—intervensi dilakukan setelah mahasiswa menunjukkan tanda-tanda akan keluar. Pendekatan berbasis data dan kecerdasan buatan menawarkan paradigma baru berupa sistem early warning yang mampu mengidentifikasi mahasiswa berisiko DO sejak dini [3]. Dalam konteks ini, machine learning (ML) telah banyak diaplikasikan pada permasalahan prediksi akademik [4]. Di antara berbagai algoritma ML, Random Forest dan XGBoost merupakan dua metode ensemble yang paling banyak digunakan. Random Forest membangun sejumlah pohon keputusan secara acak [5], sedangkan XGBoost menggunakan gradient boosting yang membangun pohon secara sekuensial [6]. Penelitian ini bertujuan: (1) membandingkan performa keduanya dalam memprediksi risiko DO; (2) mengidentifikasi fitur paling berpengaruh; dan (3) menganalisis relevansinya bagi PTI Indonesia.

2. Tinjauan Pustaka

2.1. Prediksi Drop Out dengan Machine Learning

Kumar et al. [7] membandingkan enam algoritma data mining pada dataset UCI Student Performance dan menemukan Random Forest serta XGBoost secara konsisten unggul, masing-masing mencapai akurasi 87,50% dan 84,38% pada data uji, jauh melampaui Gaussian Naive Bayes yang hanya mencapai 42,71%. Chen dan Liu [8] mengoptimalkan Random Forest menggunakan algoritma genetika (GA) dan seleksi fitur berbasis *information gain* serta *Laplacian score* untuk memprediksi performa akademik siswa, dan menemukan bahwa RF konvensional hanya mencapai akurasi 87,34%, sementara RF yang dioptimalkan mencapai 93,11%, menunjukkan bahwa optimasi hyperparameter dan seleksi fitur dapat meningkatkan akurasi secara signifikan dibandingkan model dasarnya. Di Indonesia, Susanti [9] menggunakan Decision Tree C4.5 dan Naive Bayes untuk memprediksi kelulusan siswa SMK, dengan akurasi masing-masing 80% dan 75%, menunjukkan perlunya eksplorasi algoritma yang lebih canggih untuk meningkatkan akurasi prediksi.

Penelitian terbaru memperkuat temuan tersebut sekaligus menunjukkan keterbatasannya. Lim dan Harvianto [10] membandingkan Decision Tree, Random Forest, dan XGBoost untuk prediksi kategori mahasiswa dan menemukan bahwa Random Forest mencapai performa keseluruhan tertinggi dengan generalisasi yang lebih kuat dibandingkan kedua algoritma lainnya. Namun, Daniswara dan Ratnasari [11] menunjukkan bahwa keunggulan akurasi pada pada evaluasi cross-validation konvensional tidak serta-merta menjamin ketahanan model lintas kohort, melalui skema Leave-One-Group-Out (LOGO) pada 370 data akademik dari tiga kohort berbeda, ketiga model gradient boosting yang diuji mengalami penurunan performa signifikan (F1-score rata-rata di bawah 0,51), meskipun XGBoost menunjukkan stabilitas relatif lebih baik dibanding model lain di bawah pergeseran distribusi data. Firdaus dan Sipayung [12] mengatasi keterbatasan model tunggal dengan mengembangkan arsitektur stacking ensemble yang menggabungkan Random Forest, Gradient Boosting, dan XGBoost sebagai base learners

dengan Logistic Regression sebagai meta-learner, dipadukan dengan teknik SMOTE–Tomek Links untuk mengatasi ketidakseimbangan kelas, pendekatan ini berhasil mencapai akurasi 90,91% dan ROC–AUC 95,72%, mengungguli model dasar secara individual. Ketiga studi ini menegaskan bahwa arah penelitian terkini bergeser dari sekadar membandingkan akurasi algoritma tunggal menuju evaluasi robustness lintas kohort dan pendekatan ensemble bertingkat untuk meningkatkan keandalan model dalam konteks nyata.

2.2. Random Forest

Random Forest (RF) diperkenalkan oleh Breiman [13]. RF membangun sejumlah pohon keputusan pada bootstrap sample yang berbeda dan subset fitur acak. Prediksi akhir melalui voting mayoritas. Keunggulannya meliputi ketahanan terhadap overfitting, kemampuan menangani dimensi tinggi, dan pengukuran feature importance. Dalam konteks pendidikan tinggi, RF telah banyak digunakan untuk memprediksi status penyelesaian studi mahasiswa. Nirmala et al. [14] menerapkan RF dan XGBoost dengan teknik imputasi MissForest untuk memprediksi status penyelesaian studi mahasiswa sarjana, sementara Satvika et al. [15] menunjukkan bahwa permutation feature importance dapat meningkatkan performa model k-Nearest Neighbor dalam memprediksi keberhasilan studi mahasiswa, menegaskan pentingnya interpretasi fitur di luar sekadar akurasi prediksi semata.

2.3. XGBoost

XGBoost merupakan algoritma gradient boosting yang dikembangkan oleh Chen dan Guestrin [16] Model membangun decision tree secara sekuensial, dengan setiap pohon berikutnya berupaya memperbaiki kesalahan prediksi model sebelumnya. XGBoost menggunakan regularisasi L1 dan L2 untuk mengendalikan kompleksitas model dan meningkatkan kemampuan generalisasi. Pada penelitian ini, XGBoost diimplementasikan menggunakan XGBClassifier dari library XGBoost, bukan GradientBoostingClassifier dari scikit-learn. Hyperparameter XGBoost ditentukan melalui GridSearchCV dengan kombinasi n_estimators, max_depth, learning_rate, subsample, dan colsample_bytree.

2.4. Dataset UCI Predict Students' Dropout

Dataset dikurasi oleh Realinho et al. [17] dari Polytechnic Institute of Portalegre, Portugal. Berisi 4.424 instance dengan 36 fitur mencakup informasi demografis, latar belakang akademik dan sosial-ekonomi, serta kinerja akademik. Label: Dropout, Graduate, Enrolled. Dataset ini telah menjadi benchmark internasional yang luas digunakan.

3. Metodologi Penelitian

3.1. Dataset

Dataset yang digunakan adalah *Predict Students' Dropout and Academic Success* yang tersedia melalui UCI Machine Learning Repository. Dataset dikurasi berdasarkan data mahasiswa Polytechnic Institute of Portalegre, Portugal, dan terdiri atas 4.424 instance dengan 36 fitur prediktor serta satu variabel target. Variabel prediktor mencakup karakteristik demografis, latar belakang pendidikan, kondisi sosial-ekonomi, serta performa akademik mahasiswa. Variabel target terdiri atas tiga kelas, yaitu Dropout, Graduate, dan Enrolled. Tabel 1 merangkum fitur utama yang digunakan.

Untuk penelitian ini, kelas Enrolled dikeluarkan sehingga permasalahan klasifikasi difokuskan pada dua kelas, yaitu Dropout dan Graduate. Setelah proses tersebut, dataset eksperimen terdiri atas 3.630 mahasiswa, yang terdiri atas 1.421 mahasiswa Dropout dan 2.209 mahasiswa Graduate.

Karena dataset berasal dari institusi pendidikan tinggi di Portugal, hasil penelitian ini tidak dimaksudkan untuk secara langsung merepresentasikan karakteristik mahasiswa Perguruan Tinggi Islam atau perguruan tinggi di Indonesia. Penggunaan dataset UCI dalam penelitian ini diposisikan sebagai studi eksperimental dan *benchmarking* untuk membandingkan kemampuan prediktif Random Forest dan XGBoost. Validasi menggunakan data aktual perguruan tinggi di Indonesia diperlukan untuk menguji generalisasi model pada konteks institusional, akademik, sosial, dan finansial yang berbeda.

Tabel 1. Deskripsi Fitur Dataset UCI Predict Students' Dropout

Fitur	Deskripsi	Tipe Data	Relevansi
Marital Status	Status pernikahan mahasiswa	Kategorikal	Tinggi
Application Mode	Jalur masuk ke perguruan tinggi	Kategorikal	Sedang
Course	Program studi yang diambil	Kategorikal	Tinggi
Curricular Units 2nd Sem (Approved)	Jumlah mata kuliah semester 2 yang disetujui/lulus	Numerik	Sangat Tinggi
Previous Qualification	Kualifikasi akademik sebelumnya	Kategorikal	Tinggi
Age at Enrollment	Usia saat mendaftar	Numerik	Tinggi
Scholarship Holder	Status penerima beasiswa	Biner	Sangat Tinggi
Tuition Fees Up to Date	Status pembayaran UKT tepat waktu	Biner	Sangat Tinggi
GDP	Indikator ekonomi makro	Numerik	Rendah
Target (Label)	Dropout / Graduate / Enrolled	Kategorikal	-

3.2. Preprocessing dan Validasi

Tahapan preprocessing terdiri atas pembersihan data, konversi tipe data, penanganan missing value, pembentukan label biner, normalisasi, serta penanganan ketidakseimbangan kelas. Seluruh fitur prediktor dikonversi ke format numerik sesuai kebutuhan model dan nilai yang hilang diisi menggunakan median fitur. Dataset kemudian difokuskan pada kelas Dropout dan Graduate, dengan Dropout diberi label 1 dan Graduate diberi label 0. Tabel 2 merangkum seluruh tahapan preprocessing. Untuk menghindari data leakage, seluruh tahapan preprocessing ditempatkan di dalam pipeline. MinMaxScaler dan SMOTE diterapkan pada proses pelatihan di dalam setiap fold, sehingga tidak diterapkan secara langsung pada keseluruhan dataset sebelum pembagian data. Hyperparameter tuning dilakukan dengan GridSearchCV 5-fold. Evaluasi utama kemudian dilakukan menggunakan RepeatedStratifiedKFold dengan 5 fold dan 10 pengulangan, sehingga diperoleh 50 evaluasi untuk setiap model dan setiap skenario fitur.

Selain itu, penelitian melakukan dua skenario seleksi fitur. Skenario pertama menggunakan seluruh 36 fitur. Skenario kedua menghilangkan tiga variabel makroekonomi, yaitu Unemployment Rate, Inflation Rate, dan GDP, sehingga tersisa 33 fitur. Pemilihan untuk menguji skenario tanpa variabel makroekonomi dilakukan karena ketiga variabel tersebut merupakan indikator ekonomi makro yang bersifat agregat dan tidak secara langsung menggambarkan karakteristik individual mahasiswa. Untuk memastikan bahwa penghapusan tersebut tidak sekadar didasarkan pada asumsi, dilakukan ablation study dengan membandingkan performa model menggunakan seluruh fitur dengan performa setelah ketiga variabel tersebut dihapus.

Tabel 2. Tahapan Preprocessing Data

Tahap	Proses	Keterangan
1	Seleksi Fitur	Menghapus fitur makroekonomi (GDP, Inflation Rate, Unemployment Rate) pada skenario kedua
2	Binarisasi Label	Kelas Enrolled dikeluarkan; fokus Dropout (1) vs Graduate (0)
3	Penanganan Imbalanced Data	SMOTE diterapkan pada data latih di dalam setiap fold (DO: 1.421 -> 1.767)
4	Normalisasi	MinMaxScaler untuk semua fitur numerik, di dalam pipeline
5	Validasi	RepeatedStratifiedKFold, 5 fold x 10 pengulangan (50 evaluasi per model)

3.3. Konfigurasi Model

Random Forest dan XGBoost digunakan sebagai model pembanding. Skema evaluasi performa menggunakan **nested cross-validation**, terdiri atas *outer loop* Repeated Stratified K-Fold (5 fold × 10

pengulangan, sehingga diperoleh 50 evaluasi) dan *inner loop* GridSearchCV 5-fold untuk hyperparameter tuning, dengan F1-Score sebagai fungsi objektif. Pada setiap outer fold, inner-loop GridSearchCV dijalankan ulang hanya menggunakan data latih pada fold tersebut, sehingga data uji pada setiap outer fold tidak pernah dilibatkan dalam proses tuning. Skema ini digunakan untuk menghindari kebocoran data (*data leakage*) antara proses pemilihan hyperparameter dan proses evaluasi performa.

Random Forest diuji dengan kombinasi n_estimators 100 dan 200, max_depth None, 10, dan 20, serta min_samples_split 2 dan 5. XGBoost diuji menggunakan kombinasi n_estimators 100 dan 200, max_depth 4 dan 6, learning_rate 0,05 dan 0,1, subsample 0,8 dan 1,0, serta colsample_bytree 0,8 dan 1,0.

Karena inner-loop GridSearchCV dijalankan ulang pada setiap outer fold, kombinasi hyperparameter yang terpilih berpotensi bervariasi antar-fold sesuai karakteristik data latih pada fold tersebut. Sebagai representasi konfigurasi model, dilakukan pencarian hyperparameter tambahan menggunakan GridSearchCV 5-fold pada keseluruhan dataset, terpisah dari skema nested cross-validation yang digunakan untuk evaluasi performa. Konfigurasi hasil pencarian ini digunakan untuk membentuk model akhir yang dipakai pada analisis interpretasi model (*permutation feature importance*, subbab 4.5) dan disajikan pada Tabel 3. Dengan demikian, Tabel 3 tidak merepresentasikan konfigurasi tunggal yang digunakan pada seluruh 50 evaluasi performa di Tabel 4–6, melainkan konfigurasi yang digunakan khusus untuk keperluan interpretasi fitur pada model akhir.

Tabel 3. Konfigurasi Hyperparameter Final Model

Parameter	Random Forest	XGBoost
n_estimators	200	200
max_depth	20	6
min_samples_split	2	-
learning_rate	-	0.1
subsample	-	0.8
colsample_bytree	-	0.8
random_state	42	42

3.4. Uji Statistik

Untuk menentukan apakah perbedaan performa Random Forest dan XGBoost bersifat konsisten secara statistik, digunakan Wilcoxon Signed-Rank Test dua arah pada hasil evaluasi berpasangan dari 50 outer fold nested cross-validation. Pengujian dilakukan untuk Accuracy, Precision, Recall, F1-Score, dan ROC-AUC dengan tingkat signifikansi $\alpha = 0{,}05$. Perbedaan dianggap signifikan apabila nilai $p < 0{,}05$.

3.5. Ablation Study

Ablation study dilakukan pada kedua model, yaitu Random Forest dan XGBoost, untuk mengevaluasi kontribusi tiga variabel makroekonomi, yaitu GDP, inflation rate, dan unemployment rate. Ketiga variabel ini dipilih karena merepresentasikan faktor eksternal agregat yang berbeda karakteristiknya dari fitur akademik dan demografis individual dalam dataset. Dua skenario fitur dibandingkan: (1) model dengan seluruh 36 fitur dan (2) model tanpa ketiga variabel makroekonomi tersebut, sehingga tersisa 33 fitur. Kedua skenario dievaluasi menggunakan skema nested cross-validation yang sama seperti pada subbab 3.3, sehingga diperoleh 50 evaluasi untuk setiap kombinasi model dan skenario fitur. Perbedaan nilai rata-rata Accuracy, Precision, Recall, F1-Score, dan ROC-AUC antara kedua skenario pada masing-masing model digunakan untuk menilai sensitivitas performa terhadap penghapusan variabel tersebut.

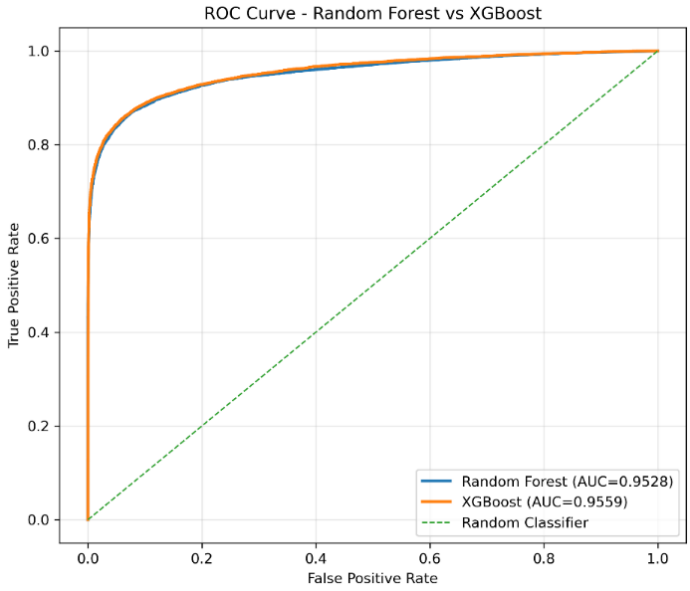
4. Hasil Dan Pembahasan

4.1. Komparasi Performa Model

Tabel 4 menyajikan hasil Repeated Stratified 5-Fold Cross-Validation (5 fold x 10 pengulangan, 50 evaluasi) untuk kedua model pada dua skenario fitur.

Tabel 4. Hasil Repeated Stratified 5-Fold Cross-Validation (5 fold x 10 pengulangan)

Model	Feature Set	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	All Features	0,9067 ± 0,0099	0,9210 ± 0,0206	0,8338 ± 0,0202	0,8750 ± 0,0134	0,9531 ± 0,0072
XGBoost	All Features	0,9085 ± 0,0103	0,9193 ± 0,0204	0,8406 ± 0,0211	0,8779 ± 0,0139	0,9568 ± 0,0071
Random Forest	Without Macroeconomic	0,9064 ± 0,0088	0,9209 ± 0,0181	0,8329 ± 0,0195	0,8745 ± 0,0121	0,9530 ± 0,0074
XGBoost	Without Macroeconomic	0,9089 ± 0,0095	0,9196 ± 0,0209	0,8415 ± 0,0188	0,8785 ± 0,0125	0,9557 ± 0,0073



Gambar 1. Kurva ROC Random Forest dan XGBoost pada skenario seluruh fitur.

Gambar 2 menyajikan kurva ROC kedua model, yang menggambarkan trade-off antara True Positive Rate dan False Positive Rate pada berbagai ambang klasifikasi. Kedua kurva berada jauh di atas garis diagonal random classifier, dengan XGBoost menghasilkan AUC 0,9559 dan Random Forest 0,9528, sejalan dengan nilai ROC-AUC pada Tabel 4. Selisih kecil antara nilai AUC pada kurva dan nilai rata-rata pada Tabel 4 (masing-masing < 0,001) wajar terjadi karena kedua nilai dihitung dengan cara berbeda: AUC pada kurva dihitung dari gabungan (pooled) seluruh prediksi out-of-fold dari 50 outer fold nested cross-validation menjadi satu kurva ROC tunggal, sedangkan nilai pada Tabel 4 merupakan rata-rata dari 50 nilai AUC yang dihitung terpisah pada masing-masing fold beserta standar deviasinya.

Hasil menunjukkan bahwa XGBoost memiliki nilai rata-rata yang sedikit lebih tinggi daripada Random Forest pada Accuracy, Recall, F1-Score, dan ROC-AUC, sedangkan Precision kedua model hampir identik. Perbedaan rata-rata relatif kecil, sehingga keunggulan numerik tidak cukup dijadikan dasar untuk menyatakan bahwa salah satu model secara mutlak lebih unggul tanpa mempertimbangkan pengujian statistik pada Bab 4.2.

4.2. Uji Statistik

Tabel 5. Hasil Uji Wilcoxon Signed-Rank

Metric	Wilcoxon Statistic	p-value	Signifikan (α=0,05)
Accuracy	439,5	0,08385	Tidak
Precision	555,0	0,43191	Tidak
Recall	152,0	0,0000607	Ya

F1-Score	413,0	0,0273	Ya
ROC-AUC	139,0	$1{,}867 \times 10^{-8}$	Ya

Uji Wilcoxon signed-rank menunjukkan adanya perbedaan yang signifikan antara Random Forest dan XGBoost pada Recall, F1-Score, dan ROC-AUC, sedangkan pada Accuracy dan Precision tidak ditemukan perbedaan yang signifikan ($p > 0{,}05$). Temuan ini menunjukkan bahwa meskipun XGBoost secara numerik unggul pada seluruh metrik yang diuji, keunggulan tersebut baru terbukti konsisten secara statistik pada metrik yang berkaitan dengan sensitivitas deteksi kelas Dropout (Recall) dan kualitas pemisahan kelas secara keseluruhan (ROC-AUC), bukan pada akurasi klasifikasi secara umum. Dengan demikian, pernyataan mengenai model yang lebih unggul perlu dibatasi pada metrik dan skema evaluasi yang digunakan. Hasil statistik ini tidak berarti bahwa XGBoost atau Random Forest selalu lebih baik pada dataset atau institusi lain.

4.3. Ablation Study

Tabel 6. Hasil Ablation Study terhadap Variabel Makroekonomi

Model	Metric	All Features	Without Macroeconomic	Difference
Random Forest	Accuracy	0,906749	0,90641	−0,000330
Random Forest	Precision	0,921040	0,920879	−0,000161
Random Forest	Recall	0,833785	0,832941	−0,000844
Random Forest	F1	0,874978	0,874466	−0,000512
Random Forest	ROC-AUC	0,953089	0,952995	−0,000094
XGBoost	Accuracy	0,908540	0,908926	+0,000386
XGBoost	Precision	0,919350	0,919634	+0,000284
XGBoost	Recall	0,840609	0,841454	+0,000845
XGBoost	F1	0,877949	0,878541	+0,000592
XGBoost	ROC-AUC	0,956754	0,955712	−0,001042

Perbandingan seluruh fitur dengan skenario tanpa fitur makroekonomi menunjukkan perubahan performa yang sangat kecil (seluruh perbedaan absolut < 0,001). Hal ini menunjukkan bahwa ketiga variabel makroekonomi tersebut tidak memberikan kontribusi prediktif yang material pada eksperimen ini. Oleh karena itu, penghapusannya dapat dipertahankan sebagai skenario reduced feature set, tetapi bukan sebagai klaim bahwa variabel tersebut secara umum tidak relevan terhadap dropout.

4.4. Analisis Confusion Matrix

Confusion matrix agregat dari Repeated Stratified 5-Fold Cross-Validation untuk kedua model disajikan pada Tabel 7a dan 7b. Dengan pengkodean label Dropout = 1 dan Graduate = 0.

Tabel 7a. Confusion Matrix Agregat Random Forest (Cross-Validation)

	Predicted Graduate	Predicted Dropout
Actual Graduate	2107	102
Actual Dropout	236	1185

Tabel 7b. Confusion Matrix Agregat XGBoost (Cross-Validation)

	Predicted Graduate	Predicted Dropout
Actual Graduate	2104	106
Actual Dropout	227	1195

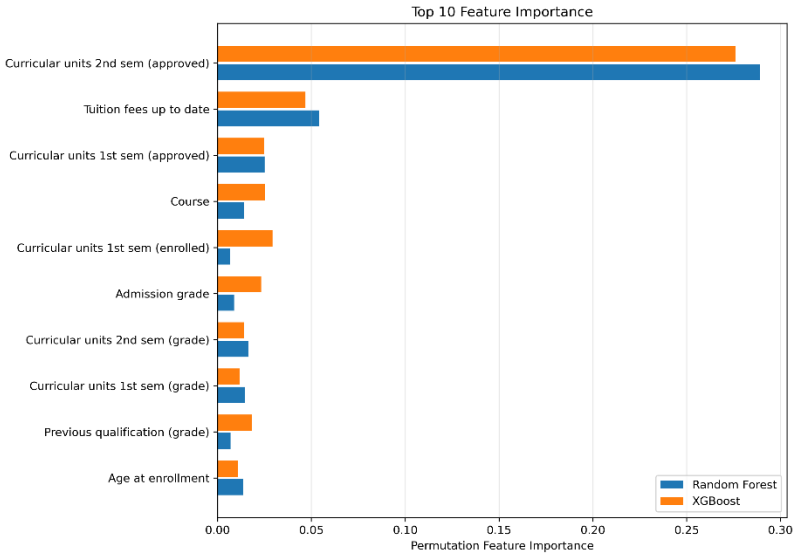
Confusion matrix pada Tabel 7a dan 7b disajikan sebagai rata-rata per repetisi dari hasil agregat *out-of-fold* Repeated Stratified 5-Fold Cross-Validation (5 fold × 10 pengulangan), dengan pengkodean label Dropout = 1 dan Graduate = 0. Berdasarkan hasil ini, Random Forest menghasilkan rata-rata 236 false negative (mahasiswa Dropout yang diprediksi Graduate) per repetisi, sedangkan XGBoost menghasilkan rata-rata 227 false negative. Dengan demikian, XGBoost menghasilkan lebih sedikit false negative dibandingkan Random Forest, sejalan dengan Recall XGBoost yang lebih tinggi dan terbukti signifikan secara statistik pada Bab 4.2. Hasil ini berbeda dari confusion matrix pada evaluasi single split (80:20) yang digunakan pada naskah sebelumnya; karena evaluasi saat ini menggunakan nested cross-validation yang lebih stabil dan bebas dari bias seleksi hyperparameter, angka pada evaluasi single split tidak lagi digunakan sebagai hasil utama.

4.5. Feature Importance

Tabel 8 menyajikan sepuluh fitur teratas berdasarkan permutation feature importance, dihitung dengan 10 pengulangan menggunakan F1-Score sebagai scoring.

Tabel 8. Top 10 Permutation Feature Importance Random Forest dan XGBoost

No	Feature	RF	SD RF	XGBoost	SD XGBoost
1	Curricular units 2nd sem (approved)	0,289087	0,005557	0,275984	0,005855
2	Tuition fees up to date	0,054092	0,002813	0,046873	0,002423
3	Curricular units 1st sem (approved)	0,025191	0,002621	0,024873	0,002392
4	Course	0,014263	0,001790	0,025484	0,001782
5	Curricular units 1st sem (enrolled)	0,006839	0,000749	0,029267	0,001064
6	Admission grade	0,008796	0,000903	0,023218	0,002060
7	Curricular units 2nd sem (grade)	0,016565	0,001337	0,014225	0,001717
8	Curricular units 1st sem (grade)	0,014654	0,001421	0,011654	0,001362
9	Previous qualification (grade)	0,006910	0,000899	0,018554	0,001421
10	Age at enrollment	0,013652	0,001664	0,011024	0,001424



Gambar 2. Top 10 permutation feature importance Random Forest dan XGBoost.

Gambar 2 memvisualisasikan sepuluh fitur teratas pada Tabel 8 dalam bentuk diagram batang berdampingan, memperlihatkan bahwa Curricular units 2nd sem (approved) memiliki kontribusi prediktif paling besar pada kedua model, dengan nilai 0,289087 untuk Random Forest dan 0,275984 untuk XGBoost. Berbeda dari metrik performa klasifikasi pada Tabel 4 dan 5 yang cenderung sedikit lebih tinggi pada XGBoost, pada aspek kontribusi fitur ini Random Forest justru menunjukkan sensitivitas permutasi yang sedikit lebih tinggi terhadap fitur akademik utama tersebut. Fitur berikutnya adalah Tuition fees up to date. Karena permutation importance mengukur perubahan performa prediksi ketika suatu fitur diacak, hasil tersebut menunjukkan kontribusi prediktif dan tidak boleh ditafsirkan sebagai hubungan kausal terhadap dropout.

4.6. Pembahasan: Interpretasi Akademik dan Finansial

Tingginya permutation importance pada Curricular units 2nd sem (approved) menunjukkan bahwa capaian akademik semester kedua merupakan sinyal prediktif yang kuat dalam dataset ini. Temuan ini sejalan dengan Romero dan Liao [18] yang juga menemukan nilai akademik semester kedua sebagai fitur terpenting pada model XGBoost dalam memprediksi dropout. Secara konseptual, capaian akademik semester awal dapat merefleksikan kemampuan mahasiswa mengikuti tuntutan pembelajaran dan mempertahankan kemajuan studi. Namun, hasil ini tidak membuktikan bahwa rendahnya capaian akademik secara kausal menyebabkan dropout.

Tuition fees up to date juga memiliki kontribusi prediktif yang relatif tinggi, sejalan dengan Umendu dkk. [19] yang melalui analisis SHAP juga mengidentifikasi status pembayaran uang kuliah sebagai salah satu prediktor paling berpengaruh terhadap hasil studi mahasiswa, bahkan pada mahasiswa dengan capaian akademik yang baik. Status pembayaran dapat menjadi indikator yang berkaitan dengan keberlanjutan studi dan kondisi finansial mahasiswa. Meskipun demikian, variabel tersebut sebaiknya dipahami sebagai indikator prediktif, bukan bukti bahwa keterlambatan pembayaran merupakan penyebab langsung dropout. Interpretasi perlu mempertimbangkan konteks institusional dan kondisi mahasiswa.

4.7. Implikasi bagi Perguruan Tinggi Islam di Indonesia

Temuan penelitian ini memberikan indikasi bahwa pendekatan *machine learning* berpotensi mendukung pengembangan sistem *early warning dropout* di perguruan tinggi. Namun, karena dataset berasal dari Polytechnic Institute of Portalegre, Portugal, model yang diperoleh belum dapat dianggap merepresentasikan mahasiswa Perguruan Tinggi Islam di Indonesia. Oleh karena itu, penerapan pada PTI Indonesia perlu melalui validasi eksternal, kalibrasi, dan pengujian pada data institusional Indonesia.

Jika telah tervalidasi, keluaran model dapat digunakan sebagai informasi pendukung untuk *early academic advising*, mentoring oleh dosen wali, konseling akademik, dan pemberian informasi mengenai dukungan finansial. Prediksi sebaiknya digunakan untuk memprioritaskan pendampingan, bukan sebagai dasar tunggal untuk menentukan status akademik mahasiswa. Pendekatan ini selaras dengan spirit hilirisasi riset dan smart campus yang menjadi tema SEMNASTIK 2026.

4.8. Pertimbangan Etis, Privasi, dan Fairness

Implementasi sistem prediksi dropout perlu memperhatikan privasi, keamanan data, fairness, dan potensi stigma terhadap mahasiswa yang dikategorikan berisiko sejalan dengan prinsip etis AI dalam pendidikan yang dirumuskan oleh Nguyen dkk [20]. Informasi prediksi harus dibatasi kepada pihak yang berwenang dan digunakan hanya untuk tujuan pendampingan pendidikan. Model tidak seharusnya menjadi dasar tunggal untuk keputusan akademik atau pemberian sanksi. Evaluasi fairness perlu dilakukan pada data institusional untuk memastikan bahwa model tidak memberikan tingkat kesalahan yang tidak proporsional pada kelompok mahasiswa tertentu, sebagaimana ditunjukkan oleh Yu dkk [21] yang mengkaji dampak penyertaan atribut terproteksi dalam model prediksi dropout terhadap kesetaraan hasil prediksi antar kelompok mahasiswa. Interpretabilitas juga penting agar pengelola dapat memahami faktor yang berkontribusi terhadap prediksi dan dapat memberikan intervensi yang sesuai.

5. Kesimpulan

Penelitian ini membandingkan performa Random Forest dan XGBoost dalam memprediksi risiko drop out mahasiswa menggunakan dataset *Predict Students' Dropout and Academic Success* dari UCI Machine Learning Repository, dengan skema nested cross-validation untuk memastikan estimasi performa yang bebas dari bias seleksi hyperparameter.
Hasil menunjukkan bahwa XGBoost memiliki nilai rata-rata sedikit lebih tinggi dibandingkan Random Forest pada seluruh metrik yang diuji, dengan Accuracy 0,9085, Recall 0,8406, F1-Score 0,8779, dan ROC-AUC 0,9568. Uji Wilcoxon Signed-Rank mengonfirmasi bahwa keunggulan XGBoost bersifat signifikan secara statistik pada Recall, F1-Score, dan ROC-AUC, tetapi tidak pada Accuracy maupun Precision. Temuan ini menunjukkan bahwa keunggulan XGBoost lebih terlihat pada kemampuannya mendeteksi mahasiswa berisiko dropout (Recall) dan memisahkan kedua kelas secara keseluruhan (ROC-AUC), dibandingkan pada akurasi klasifikasi secara umum.
Analisis permutation feature importance mengidentifikasi jumlah mata kuliah semester kedua yang disetujui (*Curricular units 2nd sem approved*) sebagai fitur paling berpengaruh pada kedua model, diikuti oleh status pembayaran UKT (*Tuition fees up to date*), menunjukkan bahwa performa akademik semester awal dan kondisi finansial mahasiswa merupakan sinyal prediktif utama terhadap risiko dropout. Sementara itu, ablation study menunjukkan bahwa penghapusan tiga variabel makroekonomi (GDP, inflation rate, dan unemployment rate) hanya menyebabkan perubahan performa yang sangat kecil (< 0,001 pada seluruh metrik), mengindikasikan bahwa faktor eksternal agregat tersebut tidak memberikan kontribusi prediktif yang material dibandingkan fitur individual mahasiswa pada dataset ini.

Penelitian ini memiliki keterbatasan penting yang perlu digarisbawahi. Dataset yang digunakan berasal dari Polytechnic Institute of Portalegre, Portugal, sehingga karakteristik mahasiswa, sistem akademik, dan konteks sosial-ekonominya berbeda dari Perguruan Tinggi Keagamaan Islam (PTKI) di Indonesia. Oleh karena itu, hasil penelitian ini bersifat eksploratif dan belum dapat digeneralisasikan secara langsung untuk memprediksi risiko dropout mahasiswa PTKI tanpa validasi eksternal menggunakan data institusional Indonesia. Meskipun demikian, pendekatan metodologis yang digunakan meliputi nested cross-validation, uji statistik, dan ablation study dapat dijadikan kerangka kerja yang dapat direplikasi pada konteks PTKI Indonesia untuk membangun sistem early warning dropout yang lebih andal. Penelitian selanjutnya perlu diarahkan pada pengumpulan dan validasi model menggunakan data aktual dari PTKI di Indonesia, disertai kajian fairness dan etika sebagaimana dibahas pada subbab 4.8, sebelum model semacam ini dapat diterapkan secara operasional untuk mendukung pendampingan akademik mahasiswa.

Referensi

[1] S. Bahri and D. M. Midyanti, “Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 1, pp. 165–172, Feb. 2023, doi: 10.25126/jtiik.2023106643.

[2] N. Wakhidah and Saemuri, “FAKTOR TERLAMBATNYA PENYELESAIAN STUDI DI PERGURUAN TINGGI KEAGAMAAN ISLAM: STUDI PENERAPAN KEBIJAKAN MASA STUDI MAHASISWA DI INDONESIA,” vol. 9, no. 4, pp. 780–802, Oct. 2025, doi: 10.1017/bpp.2023.15.

[3] F. Andika Putra, S. Mirajdandi, B. Okmarizal, and S. Mulyanda, “Prediksi Dropout Mahasiswa: Early-Warning Berbasis Enrollment dengan Machine Learning”.

[4] B. Carballo-Mendívil, A. Arellano-González, N. J. Ríos-Vázquez, and M. del P. Lizardi-Duarte, “Predicting Student Dropout from Day One: XGBoost-Based Early Warning System Using Pre-Enrollment Data,” *Applied Sciences (Switzerland)*, vol. 15, no. 16, Aug. 2025, doi: 10.3390/app15169202.

[5] L. G. R. Putra, D. D. Prasetya, and M. Mayadi, "Student Dropout Prediction Using Random Forest and XGBoost Method," *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 9, no. 1, pp. 147–157, Feb. 2025, doi: 10.29407/intensif.v9i1.21191.
[6] Abdah Syakiroh Gustian and Fathoni Mahardika, "Analisis Klasifikasi Risiko Dropout Mahasiswa Menggunakan Algoritma Decision Tree dan Random Forest," *Jupiter: Publikasi Ilmu Keteknikan Industri, Teknik Elektro dan Informatika*, vol. 3, no. 4, pp. 182–189, Jul. 2025, doi: 10.61132/jupiter.v3i4.980.
[7] M. Kumar, N. Singh, J. Wadhwa, P. Singh, G. Kumar, and A. Qtaishat, "Utilizing Random Forest and XGBoost Data Mining Algorithms for Anticipating Students' Academic Performance," *International Journal of Modern Education and Computer Science*, vol. 16, no. 2, pp. 29–44, Apr. 2024, doi: 10.5815/ijmecs.2024.02.03.
[8] M. Chen and Z. Liu, "Predicting performance of students by optimizing tree components of random forest using genetic algorithm," *Heliyon*, vol. 10, no. 12, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32570.
[9] R. Maulana, R. Narasati, R. Herdiana, R. Hamonangan, and S. Anwar, "KOMPARASI ALGORITMA DECISION TREE DAN NAIVE BAYES DALAM KLASIFIKASI PENYAKIT DIABETES," 2023.
[10] R. Calvianto Lim, "Comparative Analysis of Decision Tree, Random Forest, and XGBoost for Student Category Prediction," *International Journal of Computer Science and Humanitarian AI*, vol. 3, no. 1, pp. 9–19, 2026, doi: 10.21512/ijcshaijournal.v3i1.14936.
[11] R. Daniswara and C. I. Ratnasari, "ROBUSTNESS EVALUATION OF GRADIENT BOOSTING MODELS FOR GRADUATION PREDICTION UNDER COHORT-BASED DISTRIBUTION SHIFTS," *JIKO (Jurnal Informatika dan Komputer)*, vol. 9, no. 1, Apr. 2026, doi: 10.33387/jiko.v9i1.11546.
[12] A. N. Firdaus and Y. R. Sipayung, "Stacking Ensemble Learning for University Student Dropout Prediction," *Journal of Information Systems and Informatics*, vol. 8, no. 1, pp. 456–477, Feb. 2026, doi: 10.63158/journalisi.v8i1.1403.
[13] L. Breiman, "Random Forests," 2001.
[14] I. Nirmala, H. Wijayanto, and K. A. Notodiputro, "Prediction of Undergraduate Student's Study Completion Status Using MissForest Imputation in Random Forest and XGBoost Models," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 13, no. 1, pp. 53–62, Feb. 2022, doi: 10.21512/comtech.v13i1.7388.
[15] G. A. J. Satvika, I. N. Sukajaya, and I. G. A. Gunadi, "Improving k-nearest neighbor performance using permutation feature importance to predict student success in study," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 3, pp. 1835–1844, Sep. 2024, doi: 10.11591/ijeecs.v35.i3.pp1835-1844.
[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[17] V. Realinho, J. Machado, L. Baptista, and M. V. Martins, "Predicting Student Dropout and Academic Success," *Data (Basel).*, vol. 7, no. 11, Nov. 2022, doi: 10.3390/data7110146.
[18] S. Romero and X. Liao, "Statistical and machine learning models for predicting university dropout and scholarship impact," *PLoS One*, vol. 20, no. 6 June, Jun. 2025, doi: 10.1371/journal.pone.0325047.
[19] E. C. Umendu, M. Ghanzanfar, A. Kans, and M. A. R. Ahad, "Enhancing Student Retention in Higher Education Institutions (HEIs): Machine Learning Approach," *Electronics (Switzerland)*, vol. 15, no. 4, Feb. 2026, doi: 10.3390/electronics15040734.
[20] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B. P. T. Nguyen, "Ethical principles for artificial intelligence in education," *Educ. Inf. Technol. (Dordr).*, vol. 28, no. 4, pp. 4221–4241, Apr. 2023, doi: 10.1007/s10639-022-11316-w.
[21] R. Yu, H. Lee, and R. F. Kizilcec, "Should College Dropout Prediction Models Include Protected Attributes?," in *L@S 2021 - Proceedings of the 8th ACM Conference on Learning @*

Scale, Association for Computing Machinery, Inc, Jun. 2021, pp. 91–100. doi: 10.1145/3430895.3460139.