Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

Identifikasi Faktor Dominan Kegagalan Akademik pada Data Tidak Seimbang Menggunaan Ensemble Learning dan Hybrid SMOTE-ENN

Rizky Nurhasanah[1,*], Solikhun[2]

[1]Sistem Informasi, STIKOM Tunas Bangsa, Pematang Siantar, Indonesia
Jl. Sudirman No.1, 2 & 3, Banjar, Kec. Siantar Bar., Kota Pematang Siantar, Sumatera Utara, Indonesia
[2]Teknik Informatika, STIKOM Tunas Bangsa, Pematang Siantar, Indonesia
Jl. Sudirman No.1, 2 & 3, Banjar, Kec. Siantar Bar., Kota Pematang Siantar, Sumatera Utara, Indonesia
Email: [1,*]rizkinurhasanah733@gmail.com, [2]solikhun@amiktunasbangsa.ac.id
Email Penulis Korespondensi: Email Author Korespondensi
Submitted: **25/02/2026**; Accepted: **30/04/2026**; Published: **30/04/2026**

Abstrak−Kegagalan akademik dan dropout mahasiswa merupakan permasalahan yang berdampak pada kualitas lulusan serta efektivitas penyelenggaraan pendidikan tinggi. Salah satu tantangan dalam membangun model prediksi dropout adalah ketidakseimbangan distribusi kelas (class imbalance), yang menyebabkan model cenderung lebih akurat dalam mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas. Penelitian ini bertujuan mengidentifikasi faktor-faktor dominan yang memengaruhi kegagalan akademik dengan menerapkan teknik Hybrid Synthetic Minority Oversampling Technique–Edited Nearest Neighbour (SMOTE-ENN) yang dikombinasikan dengan algoritma Ensemble Learning, yaitu Decision Tree, Random Forest, XGBoost, dan Voting Ensemble. Hybrid SMOTE-ENN dipilih karena mampu meningkatkan representasi kelas minoritas melalui proses oversampling menggunakan SMOTE sekaligus mengurangi noise dan overlapping data menggunakan Edited Nearest Neighbour (ENN), sehingga menghasilkan distribusi data latih yang lebih seimbang. Dataset yang digunakan adalah Predict Students' Dropout and Academic Success yang terdiri atas 4.424 data. Tahapan penelitian meliputi data preprocessing, pembagian data latih dan data uji, penyeimbangan data menggunakan Hybrid SMOTE-ENN, pelatihan model, evaluasi menggunakan accuracy, precision, recall, dan F1-score, serta analisis feature importance. Hasil eksperimen menunjukkan bahwa model XGBoost dengan Hybrid SMOTE-ENN memberikan kinerja terbaik dengan accuracy sebesar 87,12%, precision 87,20%, recall 87,12%, dan F1-score 87,15%. Selain itu, model memperoleh recall sebesar 80,99% pada kelas Dropout, yang menunjukkan kemampuan yang baik dalam mengidentifikasi mahasiswa berisiko mengalami kegagalan akademik. Analisis feature importance menunjukkan bahwa Curricular Units 2nd Semester (Approved), Curricular Units 1st Semester (Approved), dan Curricular Units 2nd Semester (Grade) merupakan tiga faktor yang paling berpengaruh terhadap risiko kegagalan akademik. Hasil penelitian ini memberikan kontribusi dalam pengembangan sistem deteksi dini berbasis machine learning untuk mendukung pengambilan keputusan akademik pada data mahasiswa yang tidak seimbang.

Kata Kunci: Kegagalan Akademik; Student Dropout; Ensemble Learning, Hybrid SMOTE-ENN; Machine Learning.

Abstract−Academic failure and student dropout remain significant challenges in higher education because they affect educational quality and institutional performance. One of the major challenges in developing an effective dropout prediction model is class imbalance, which causes classification algorithms to be biased toward the majority class and reduces their ability to identify minority-class instances. This study aims to identify the dominant factors influencing academic failure by integrating the Hybrid Synthetic Minority Oversampling Technique–Edited Nearest Neighbour (SMOTE-ENN) with Ensemble Learning algorithms, including Decision Tree, Random Forest, XGBoost, and Voting Ensemble. Hybrid SMOTE-ENN was selected because it combines minority-class oversampling with noise and overlapping data removal, resulting in a more balanced training dataset and improving classification performance. The experiment was conducted using the Predict Students' Dropout and Academic Success dataset containing 4,424 student records. The research procedure consisted of data preprocessing, train–test splitting, Hybrid SMOTE-ENN resampling, model training, performance evaluation using accuracy, precision, recall, and F1-score, followed by feature importance analysis. Experimental results demonstrate that XGBoost with Hybrid SMOTE-ENN achieved the best performance, obtaining an accuracy of 87.12%, precision of 87.20%, recall of 87.12%, and F1-score of 87.15%. More importantly, the proposed model achieved a dropout-class recall of 80.99%, indicating its effectiveness in identifying students at risk of academic failure. Feature importance analysis revealed that Curricular Units 2nd Semester (Approved), Curricular Units 1st Semester (Approved), and Curricular Units 2nd Semester (Grade) are the three most influential factors affecting student dropout risk. These findings contribute to the development of an early warning system based on machine learning to support academic decision-making for imbalanced educational datasets.

Keywords: Academic Failure; Student Dropout; Ensemble Learning; Hybrid SMOTE-ENN; Machine Learning.

1. PENDAHULUAN

Kegagalan akademik dan dropout mahasiswa masih menjadi salah satu tantangan utama dalam penyelenggaraan pendidikan tinggi karena berdampak terhadap kualitas lulusan, efektivitas proses pembelajaran, akreditasi program studi, serta efisiensi pengelolaan institusi pendidikan[1], [2], [3]. Mahasiswa yang mengalami kegagalan akademik cenderung memiliki prestasi belajar yang rendah, keterlambatan masa studi, hingga memutuskan untuk berhenti kuliah (dropout), sehingga menimbulkan kerugian baik bagi mahasiswa maupun perguruan tinggi[4], [5], [6]. Oleh karena itu, identifikasi dini terhadap mahasiswa yang berpotensi mengalami kegagalan akademik menjadi langkah penting untuk mendukung penyusunan strategi pendampingan akademik dan pengambilan keputusan yang lebih tepat sasaran[7], [8], [9]. Perkembangan teknologi informasi serta ketersediaan data akademik yang semakin besar mendorong pemanfaatan Educational Data Mining (EDM) dan Machine Learning sebagai pendekatan yang efektif

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

dalam membangun sistem prediksi risiko kegagalan akademik berdasarkan pola historis data mahasiswa[10], [11], [12].

Berbagai penelitian telah menunjukkan bahwa algoritma machine learning mampu memberikan performa yang baik dalam memprediksi status akademik mahasiswa. Putra dkk. menerapkan algoritma Random Forest dan XGBoost untuk memprediksi student dropout dan menunjukkan bahwa kedua algoritma tersebut memiliki kemampuan yang baik dalam melakukan klasifikasi mahasiswa berdasarkan data akademik. Hasil penelitian tersebut menunjukkan bahwa pendekatan ensemble learning mampu meningkatkan stabilitas model dibandingkan algoritma klasifikasi tunggal [13]. Selain itu, Ricky dkk. mengembangkan model hibrida menggunakan teknik stacking dan blending pada prediksi student dropout di lingkungan Massive Open Online Course (MOOC). Hasil penelitian menunjukkan bahwa kombinasi beberapa model pembelajaran mampu meningkatkan performa prediksi dibandingkan penggunaan model tunggal [14].

Selain pendekatan ensemble learning, berbagai penelitian juga menerapkan algoritma klasifikasi pada permasalahan akademik lainnya. Hasanah dkk. membandingkan algoritma CART dan Random Forest dalam menganalisis status mahasiswa di Universitas Terbuka dan melaporkan bahwa Random Forest menghasilkan performa klasifikasi yang lebih baik dibandingkan CART [15]. Purnama dkk. menerapkan algoritma Decision Tree dan Support Vector Machine (SVM) dalam mengklasifikasikan mahasiswa yang mengulang mata kuliah (HER), serta menunjukkan bahwa metode klasifikasi mampu mendukung proses pengambilan keputusan akademik[7]. Di sisi lain, Nababan dkk. mengombinasikan algoritma XGBoost dengan teknik Synthetic Minority Oversampling Technique (SMOTE) untuk meningkatkan kemampuan model dalam mengenali kelas minoritas [16]. Hasil yang serupa juga dilaporkan oleh Anggrawan dkk., yang menunjukkan bahwa penerapan SMOTE mampu meningkatkan performa klasifikasi pada data kelulusan mahasiswa yang tidak seimbang[17]. Selain itu, Desnelita dkk. membuktikan bahwa kombinasi Random Forest dan SMOTE mampu meningkatkan performa klasifikasi dibandingkan model tanpa proses penyeimbangan data [11]. Secara umum, penelitian-penelitian tersebut menunjukkan bahwa algoritma ensemble learning dan teknik resampling memiliki potensi yang baik dalam meningkatkan performa klasifikasi pada data akademik yang tidak seimbang.

Meskipun demikian, salah satu tantangan utama dalam pengembangan model prediksi kegagalan akademik adalah ketidakseimbangan distribusi kelas (class imbalance). Pada dataset akademik, jumlah mahasiswa yang berhasil menyelesaikan studi umumnya jauh lebih besar dibandingkan mahasiswa yang mengalami dropout, sehingga model klasifikasi cenderung mempelajari karakteristik kelas mayoritas dan mengabaikan pola pada kelas minoritas. Akibatnya, model dapat menghasilkan nilai accuracy yang tinggi, tetapi memiliki kemampuan yang kurang optimal dalam mengidentifikasi mahasiswa yang benar-benar berisiko mengalami dropout. Oleh karena itu, evaluasi model pada kasus imbalanced dataset tidak cukup hanya menggunakan accuracy, tetapi juga perlu mempertimbangkan precision, recall, dan F1-score, khususnya pada kelas minoritas. Berbagai penelitian telah mengembangkan pendekatan untuk mengatasi permasalahan tersebut, namun masih terdapat beberapa keterbatasan yang membuka peluang bagi pengembangan penelitian lebih lanjut.

Berdasarkan hasil telah penelitian terdahulu, dapat disimpulkan bahwa pendekatan machine learning untuk prediksi student dropout telah berkembang melalui penerapan algoritma ensemble learning, teknik stacking, blending, serta metode resampling seperti SMOTE. Meskipun berbagai pendekatan tersebut mampu meningkatkan performa klasifikasi, masih terdapat beberapa keterbatasan. Sebagian besar penelitian hanya berfokus pada perbandingan algoritma klasifikasi atau penerapan teknik resampling secara terpisah, sehingga belum mengeksplorasi integrasi teknik Hybrid SMOTE-ENN dengan beberapa algoritma ensemble learning dalam satu kerangka eksperimen. Selain itu, penelitian terdahulu umumnya lebih menitikberatkan pada peningkatan nilai accuracy, sedangkan analisis terhadap kemampuan model dalam mendeteksi kelas minoritas (dropout) melalui metrik recall serta identifikasi faktor-faktor dominan menggunakan feature importance masih relatif terbatas. Kondisi tersebut menunjukkan adanya peluang pengembangan model yang tidak hanya mampu meningkatkan performa klasifikasi pada data akademik yang tidak seimbang, tetapi juga menghasilkan interpretasi mengenai faktor-faktor yang berkontribusi terhadap risiko kegagalan akademik.

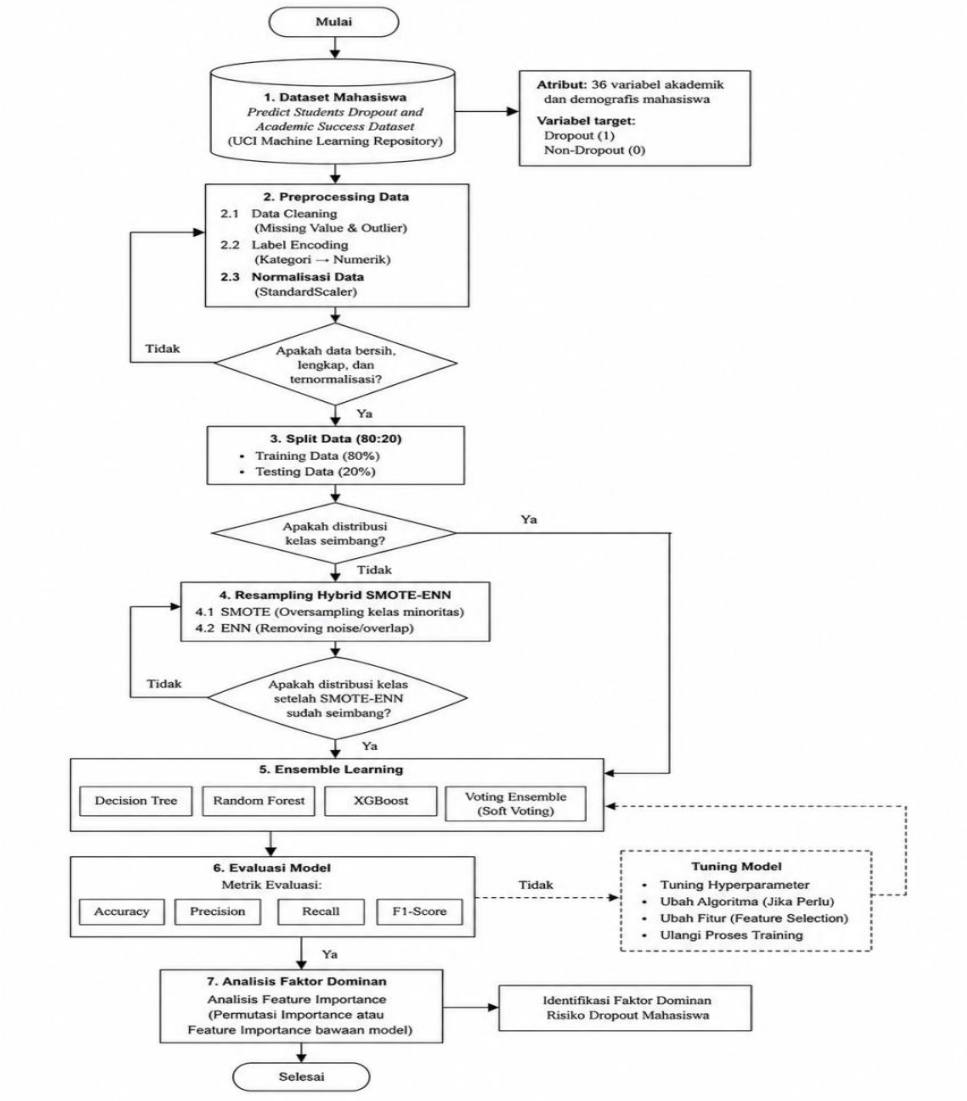
Berdasarkan research gap tersebut, penelitian ini bertujuan mengembangkan model prediksi kegagalan akademik mahasiswa menggunakan teknik Hybrid Synthetic Minority Oversampling Technique–Edited Nearest Neighbour (SMOTE-ENN) yang dikombinasikan dengan algoritma Decision Tree, Random Forest, XGBoost, dan Voting Ensemble untuk mengatasi permasalahan ketidakseimbangan distribusi kelas pada prediksi student dropout. Berbeda dengan penelitian sebelumnya, penelitian ini mengevaluasi performa model menggunakan metrik accuracy, precision, recall, dan F1-score, dengan penekanan khusus pada kemampuan model dalam mengidentifikasi kelas minoritas (dropout). Selain itu, penelitian ini menerapkan analisis feature importance untuk mengidentifikasi faktor-faktor akademik yang paling dominan memengaruhi risiko kegagalan akademik. Penelitian ini diharapkan memberikan kontribusi dalam pengembangan model prediksi student dropout yang lebih efektif pada data akademik yang tidak seimbang melalui integrasi Hybrid SMOTE-ENN dan algoritma ensemble learning. Selain itu, hasil analisis feature importance diharapkan mampu memberikan informasi mengenai faktor-faktor akademik dominan yang dapat dimanfaatkan sebagai dasar pengembangan sistem deteksi dini, strategi pendampingan akademik, serta pengambilan keputusan di lingkungan perguruan tinggi.

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan eksperimen (experimental research) berbasis machine learning untuk membangun model prediksi kegagalan akademik mahasiswa menggunakan algoritma ensemble learning dan teknik Hybrid Synthetic Minority Oversampling Technique–Edited Nearest Neighbour (SMOTE-ENN). Tahapan penelitian disusun secara sistematis mulai dari pengumpulan dataset, data preprocessing, pembagian data, penyeimbangan distribusi kelas, pelatihan model, evaluasi performa, hingga analisis feature importance. Alur metodologi penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir metodologi penelitian

Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, penelitian diawali dengan penggunaan Predict Students Dropout and Academic Success Dataset yang diperoleh dari UCI Machine Learning Repository. Dataset terdiri atas 4.424 data mahasiswa dengan 36 atribut akademik dan demografis, serta memiliki target klasifikasi Dropout dan Non-Dropout. Selanjutnya dilakukan tahap preprocessing yang meliputi data cleaning, label encoding, dan normalisasi menggunakan StandardScaler. Setelah data dinyatakan bersih dan siap digunakan, dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%). Distribusi kelas pada data pelatihan kemudian diperiksa, dan apabila masih tidak seimbang diterapkan teknik Hybrid SMOTE-ENN untuk melakukan oversampling pada kelas minoritas sekaligus menghilangkan noise menggunakan metode ENN. Dataset hasil resampling selanjutnya digunakan untuk melatih model Decision Tree, Random Forest, XGBoost, dan Voting Ensemble. Performa setiap model dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Apabila performa model belum optimal, dilakukan proses hyperparameter tuning dan pelatihan ulang. Tahap terakhir adalah analisis feature importance untuk mengidentifikasi faktor-faktor akademik yang paling dominan memengaruhi risiko dropout mahasiswa.

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

2.2 Dataset Penelitian

Penelitian ini menggunakan Predict Students Dropout and Academic Success Dataset yang diperoleh dari UCI Machine Learning Repository. Dataset terdiri atas 4.424 data mahasiswa dengan 36 atribut yang merepresentasikan karakteristik akademik, sosial, dan demografis mahasiswa[18]. Target klasifikasi terdiri atas dua kelas, yaitu Dropout dan Non-Dropout. Kelas Non-Dropout diperoleh dengan menggabungkan kelas Graduate dan Enrolled, sehingga permasalahan penelitian diformulasikan sebagai klasifikasi biner. Karakteristik dataset yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Deskripsi atribut pada Predict Students Dropout and Academic Success Dataset

Kelompok Atribut	Nama Atribut
Karakteristik Demografis	Marital status, Gender, Age at enrollment, Nationality, International
Latar Belakang Pendidikan	Previous qualification, Previous qualification (grade), Admission grade, Educational special needs
Kondisi Sosial Ekonomi	Debtor, Tuition fees up to date, Scholarship holder, Displaced
Karakteristik Orang Tua	Mother's qualification, Father's qualification, Mother's occupation, Father's occupation
Informasi Akademik	Application mode, Application order, Course, Daytime/Evening attendance
Performa Akademik Semester 1	Curricular units 1st sem (credited), enrolled, evaluations, approved, grade, without evaluations
Performa Akademik Semester 2	Curricular units 2nd sem (credited), enrolled, evaluations, approved, grade, without evaluations
Indikator Makroekonomi	Unemployment rate, Inflation rate, GDP
Variabel Target	Dropout, Graduate, Enrolled (diubah menjadi Dropout dan Non-Dropout)

Target klasifikasi terdiri dari dua kelas yaitu Dropout dan Non-Dropout. Kelas Non-Dropout merupakan gabungan dari kelas Graduate dan Enrolled. Permasalahan klasifikasi pada penelitian ini dapat direpresentasikan menggunakan Persamaan (1).

$$f(x_i) \rightarrow y_i \tag{1}$$

2.3 Preprocessing Data

Tahap preprocessing dilakukan untuk menghasilkan data yang bersih dan siap digunakan pada proses pelatihan model. Tahapan ini meliputi data cleaning, label encoding, dan normalisasi menggunakan StandardScalerv [19]. Label encoding mengubah atribut kategorikal menjadi numerik, sedangkan StandardScaler menyeragamkan skala atribut dengan rata-rata (mean) 0 dan simpangan baku (standard deviation) 1 sebagaimana ditunjukkan pada Persamaan (2).

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

2.4 SMOTE-ENN

Dataset penelitian memiliki distribusi kelas tidak seimbang sehingga digunakan teknik hybrid resampling SMOTE-ENN. Metode SMOTE digunakan untuk menghasilkan data sintetis pada kelas minoritas menggunakan pendekatan k-nearest neighbors [9]-[24]. Pembentukan data sintetis dilakukan menggunakan Persamaan (3).

$$x_{baru} = x_i + \delta(x_{zi} - x_i) \tag{3}$$

Setelah proses SMOTE dilakukan, metode Edited Nearest Neighbour (ENN) digunakan untuk menghapus data noise dan data ambigu pada batas keputusan kelas sehingga kualitas distribusi data menjadi lebih baik [25]. Pada metode ENN, data akan dihapus apabila mayoritas tetangga terdekat memiliki label kelas yang berbeda. Proses ENN dapat direpresentasikan menggunakan Persamaan (4).

$$ENN(x_i) = \begin{cases} \text{hapus}, & \text{jika majority}(N\ N_k(x_i) \neq y_i) \\ \text{simpan}, & \text{jika majority}(N\ N_k(x_i) = y_i) \end{cases} \tag{4}$$

2.5 Ensemble Learning

Metode ensemble learning merupakan pendekatan machine learning yang menggabungkan beberapa model klasifikasi untuk meningkatkan performa prediksi dan menghasilkan model yang lebih stabil dibandingkan model tunggal. Pendekatan ini mampu mengurangi overfitting, meningkatkan akurasi klasifikasi, serta bekerja lebih baik pada data kompleks dan tidak seimbang[26]. Selain itu, beberapa algoritma ensemble memiliki kemampuan dalam menghasilkan feature importance sehingga dapat digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil prediksi. Algoritma ensemble learning yang digunakan terdiri dari Decision Tree, Random Forest, XGBoost, dan Voting Ensemble sebagai berikut:

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

2.5.1 Decision Tree

Decision Tree digunakan sebagai model dasar klasifikasi berbasis pohon keputusan[27], [28]. Algoritma ini membagi data berdasarkan atribut terbaik menggunakan nilai entropy dan information gain. Perhitungan entropy ditunjukkan pada Persamaan (5).

$$\text{Entropy}(S) = \sum_{i=1}^{n} p_i \log_2 p_i \quad (5)$$

Sedangkan Information gain dihitung menggunaka Persamaan (6).

$$\text{Gain}(S,\ A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(s_v) \quad (6)$$

2.5.2 Random Forest

Random Forest bekerja menggunakan teknik bagging dengan banyak pohon keputusan untuk meningkatkan stabilitas model dan mengurangi overfitting[29]. Prediksi akhir Random Forest diperoleh berdasarkan voting mayoritas dari seluruh pohon keputusan seperti pada Persamaan (7).

$$RF(x) = \text{mode}(T_1(x), T_2(x), \cdots, T_n(x)) \quad (7)$$

2.5.3 XGBoost

XGBoost mengimplementasikan metode gradient boosting melalui proses pembelajaran bertahap, di mana setiap iterasi dirancang untuk memperbaiki kesalahan prediksi dari model sebelumnya sehingga menghasilkan kinerja klasifikasi yang lebih optimal [30]. Fungsi objektif XGBoost dapat direpresentasikan menggunakan Persamaan (8).

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} l(y_i,\ \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (8)$$

Regularisasi digunakan untuk mengurangi kompleksitas model dan mencegah overfitting.

2.5.4 Voting Ensemble

Voting Ensemble d|igunakan untuk menggabungkan hasil prediksi beberapa model menggunakan metode hard voting[31]. Prediksi akhir ditentukan berdasarkan suara mayoritas dari seluruh model klasifikasi seperti pada Persamaan (9).

$$V(x) = \text{mode}(h_1(x), h_2(x), \cdots, h_n(x)) \quad (9)$$

2.6 Evaluation Metrics

Evaluasi model dilakukan menggunakan confusion matrix dan beberapa metrik evaluasi yaitu accuracy, precision, recall, dan F1-score[32]. Persamaan evaluasi ditunjukkan pada Persamaan (10)–(13).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

$$F1 - \text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precesion} + \text{Recall}} \quad (13)$$

Selain evaluasi performa model, penelitian ini juga melakukan analisis feature importance menggunakan Random Forest dan XGBoostuntuk menentukan faktor dominan kegagalan akademik mahasiswa.

3. HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan hasil pengujian model machine learning menggunakan metode Decision Tree, Random Forest, XGBoost, dan Voting Ensemble dengan teknik hybrid resampling SMOTE-ENN untuk mengidentifikasi faktor dominan kegagalan akademik mahasiswa. Hasil penelitian meliputi distribusi dataset, proses resampling, evaluasi performa model, confusion matrix, cross-validation, serta analisis feature importance.

3.1 Hasil Dataset dan Recording

3.1.1 Distribusi Dataset

Dataset yang digunakan terdiri atas 4.424 data mahasiswa dengan 36 atribut prediktor dan 1 variabel target, sehingga keseluruhan dataset memiliki 37 kolom. Dataset memiliki distribusi kelas yang tidak seimbang, di mana kelas Graduate memiliki jumlah data paling besar dibandingkan kelas Dropout dan Enrolled. Kondisi ini dapat menyebabkan model menjadi bias terhadap kelas mayoritas.

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

Tabel 2. Distribusi Dataset Awal

Kelas	Jumlah Data	Persentase
Graduate	2.209	49,9%
Dropout	1.421	32,1%
Enrolled	794	17,9%

Berdasarkan Tabel 2, kelas Graduate memiliki jumlah data terbesar yaitu 2.209 data (49,9%), sedangkan kelas Enrolled merupakan kelas dengan jumlah data paling sedikit yaitu 794 data (17,9%). Perbedaan distribusi antar kelas menunjukkan bahwa dataset berada dalam kondisi imbalanced, sehingga berpotensi menyebabkan model lebih dominan mempelajari karakteristik kelas mayoritas.

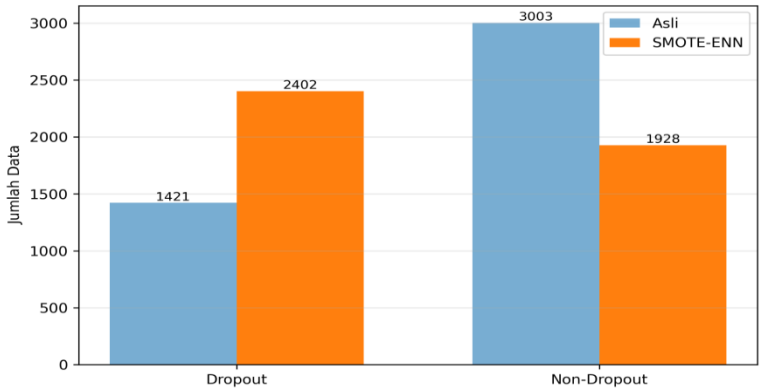
Tabel 3. Distribusi Target Klasifikasi

Target	Jumlah Data	Persentase
Non-Dropout	3.003	67,90%
Dropout	1.421	32,10%

Berdasarkan Tabel 3, setelah kelas Graduate dan Enrolled digabung menjadi Non-Dropout, diperoleh distribusi kelas sebesar 67,9% untuk kelas Non-Dropout dan 32,1% untuk kelas Dropout. Ketidakseimbangan distribusi tersebut berpotensi menyebabkan model menghasilkan bias terhadap kelas mayoritas, sehingga diperlukan teknik Hybrid SMOTE-ENN untuk menyeimbangkan distribusi data pelatihan.

3.1.2 Distribusi Dataset

Setelah pembagian data dengan rasio 80% data pelatihan dan 20% data pengujian, teknik Hybrid SMOTE-ENN diterapkan pada data pelatihan untuk mengatasi ketidakseimbangan distribusi kelas. Sebelum proses resampling, data pelatihan berjumlah 3.539 data, kemudian meningkat menjadi 4.391 data setelah penerapan Hybrid SMOTE-ENN. Perubahan distribusi kelas sebelum dan sesudah proses resampling disajikan pada Gambar 2.



Gambar 2. Distribusi Kelas Sebelum dan Sesudah Penerapan Hybrid SMOTE-ENN

Gambar 2 menunjukkan perubahan distribusi kelas sebelum dan sesudah penerapan Hybrid SMOTE-ENN. Sebelum proses resampling, distribusi data didominasi oleh kelas Non-Dropout sebanyak 3.003 data, sedangkan kelas Dropout hanya berjumlah 1.421 data. Setelah penerapan Hybrid SMOTE-ENN, jumlah data pada kelas Dropout meningkat menjadi 2.402 data, sedangkan kelas Non-Dropout menjadi 1.989 data. Hasil tersebut menunjukkan bahwa Hybrid SMOTE-ENN berhasil menghasilkan distribusi kelas yang lebih seimbang sehingga mengurangi dominasi kelas mayoritas. Kondisi ini memberikan peluang bagi model klasifikasi untuk mempelajari karakteristik kelas minoritas secara lebih baik dan meningkatkan kemampuan dalam mendeteksi mahasiswa yang berisiko mengalami dropout .

3.2 Hasil Evaluasi Model Klasifikasi

3.2.1 Hasil Evaluasi Model

Pengujian dilakukan terhadap algoritma Decision Tree, Random Forest, XGBoost, dan Voting Ensemble. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score untuk membandingkan performa model sebelum dan sesudah penerapan Hybrid SMOTE-ENN.

Tabel 4. Hasil Perbandingan Evaluasi Model Setelah SMOTE-ENN

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree Baseline	81,13%	81,30%	81,13%	81,21%
Random Forest Baseline	87,12%	87,11%	87,12%	86,72%

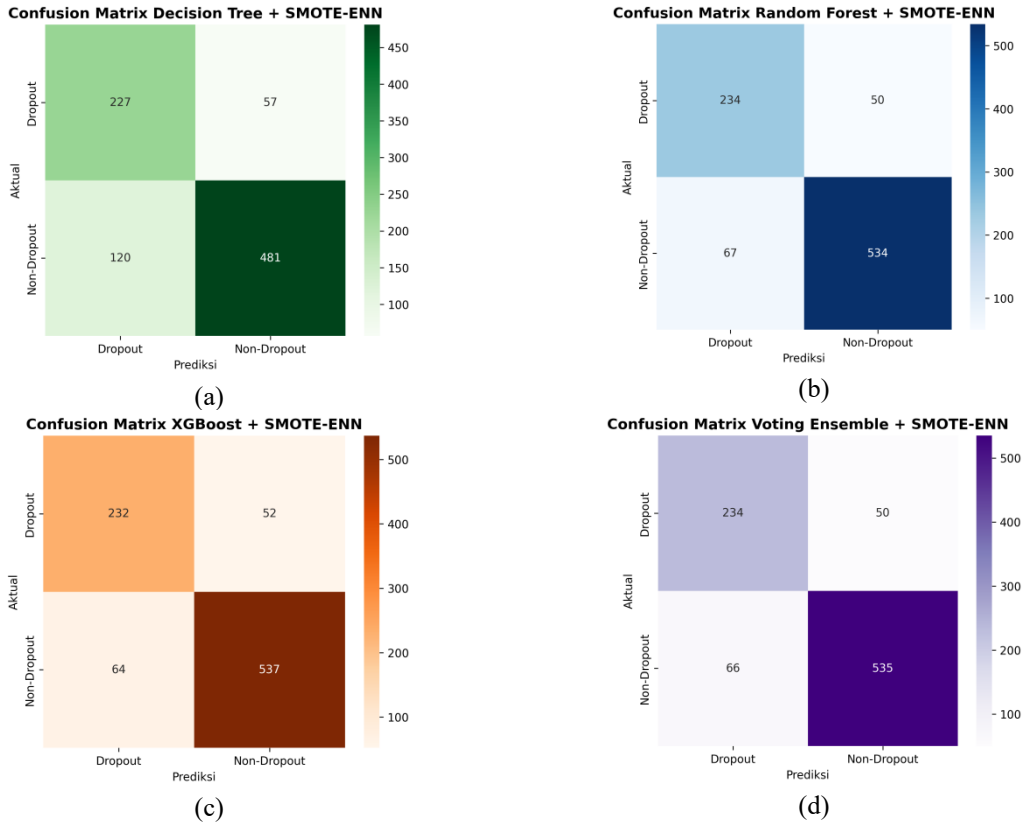
Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

Model	Accuracy	Precision	Recall	F1-Score
XGBoost Baseline	86,33%	86,19%	86,33%	85,97%
Decision Tree + SMOTE-ENN	79,21%	80,84%	79,21%	79,66%
Random Forest + SMOTE-ENN	87,01%	87,19%	87,01%	87,08%
XGBoost + SMOTE-ENN	87,12%	87,20%	87,12%	87,15%
Voting Ensemble + SMOTE-ENN	86,89%	87,00%	86,89%	86,94%

Berdasarkan Tabel 5, model XGBoost + Hybrid SMOTE-ENN memperoleh performa terbaik dengan accuracy sebesar 87,12%, precision 87,20%, recall 87,12%, dan F1-score 87,15%. Dibandingkan model baseline, penerapan Hybrid SMOTE-ENN memberikan peningkatan pada Random Forest dan XGBoost, sedangkan Voting Ensemble juga menunjukkan performa yang kompetitif dengan accuracy sebesar 86,89%. Sebaliknya, Decision Tree mengalami penurunan performa setelah proses resampling, yang menunjukkan bahwa tidak semua algoritma memperoleh peningkatan yang sama akibat penerapan Hybrid SMOTE-ENN. Secara keseluruhan, hasil tersebut menunjukkan bahwa kombinasi XGBoost dan Hybrid SMOTE-ENN memberikan keseimbangan performa terbaik dalam mengklasifikasikan mahasiswa yang berisiko mengalami dropout pada dataset yang tidak seimbang.

3.2.2 Hasil Confusion Matrix

Confusion matrix digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan kelas Dropout dan Non-Dropout berdasarkan jumlah prediksi benar maupun kesalahan klasifikasi[27]. Hasil confusion matrix untuk setiap model setelah penerapan Hybrid SMOTE-ENN ditunjukkan pada Gambar 4.



Gambar 4. Confusion Matrix (a) Decision Tree, (b) Random Forest, (c) XGBoost, dan (d)Voring EnsembleSeterlah Penerapan Hybrid SMOTE-EEN

Berdasarkan Gambar 4, seluruh model mampu mengklasifikasikan kedua kelas dengan baik setelah proses Hybrid SMOTE-ENN. Model Decision Tree menghasilkan jumlah kesalahan klasifikasi yang lebih tinggi dibandingkan model lainnya, sedangkan Random Forest, XGBoost, dan Voting Ensemble menunjukkan jumlah prediksi benar yang lebih besar. Di antara seluruh model, XGBoost memberikan performa terbaik dengan 232 prediksi benar pada kelas Dropout dan 537 prediksi benar pada kelas Non-Dropout, yang sejalan dengan hasil evaluasi model sebelumnya. Hasil tersebut menunjukkan bahwa penerapan Hybrid SMOTE-ENN mampu meningkatkan kemampuan model dalam mendeteksi mahasiswa yang berisiko mengalami dropout pada dataset yang tidak seimbang.

3.2.3 Classification Report

Kinerja klasifikasi pada setiap kelas target dievaluasi menggunakan classification report berdasarkan metrik precision, recall, dan F1-score.

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

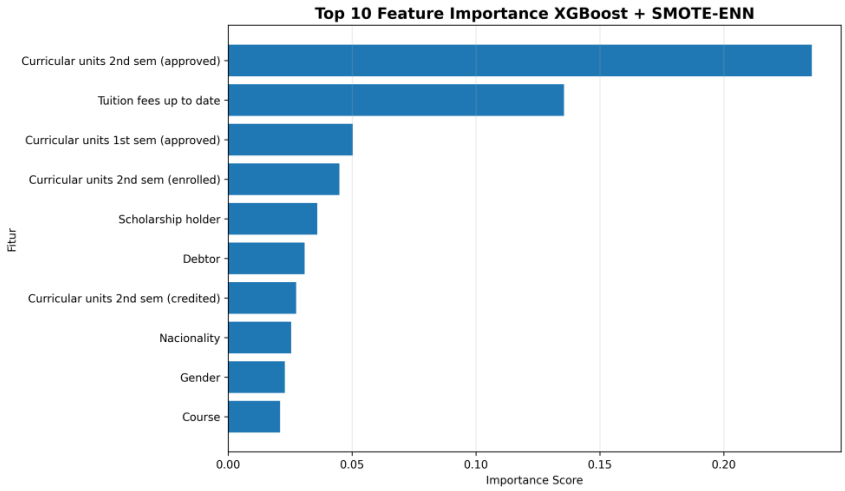
Tabel 5. Classification Report XGBoost SMOTE-ENN

Model	Kelas	Precision	Recall	F1-Score
Decision Tree	Dropout	64,53%	78,17%	70,70%
	Non-Dropout	88,54%	79,70%	83,89%
Random Forest	Dropout	78,45%	82,04%	80,21%
	Non-Dropout	91,33%	89,35%	90,33%
XGBoost	Dropout	79,31%	80,99%	80,14%
	Non-Dropout	90,92%	89,35%	90,47%
Voting Ensemble	Dropout	78,77%	80,99%	79,86%
	Non-Dropout	90,89%	89,02%	90,28%

Kinerja klasifikasi setiap model dievaluasi menggunakan classification report berdasarkan metrik precision, recall, dan F1-score. Hasil pada Tabel 6 menunjukkan bahwa Decision Tree menghasilkan performa terendah, terutama pada kelas Dropout, dengan F1-score sebesar 70,70%. Random Forest, XGBoost, dan Voting Ensemble memberikan performa yang lebih baik setelah penerapan Hybrid SMOTE-ENN. Di antara seluruh model, XGBoost memperoleh hasil paling seimbang dengan precision 79,31%, recall 80,99%, dan F1-score 80,14% pada kelas Dropout, serta F1-score 90,47% pada kelas Non-Dropout. Hasil tersebut menunjukkan bahwa penerapan Hybrid SMOTE-ENN mampu meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa menurunkan performa klasifikasi pada kelas mayoritas secara signifikan.

3.3 Hasil Feature Impoertance

Setelah diperoleh model terbaik, yaitu XGBoost dengan Hybrid SMOTE-ENN, dilakukan analisis feature importance untuk mengidentifikasi atribut yang memberikan kontribusi terbesar terhadap proses prediksi dropout. Analisis ini dilakukan untuk mengetahui faktor-faktor yang paling berpengaruh dalam pengambilan keputusan model. Hasil feature importance ditunjukkan pada Gambar 5.



Gambar 5. Feature Importance XGBoost

Berdasarkan Gambar 5, atribut Curricular units 2nd semester (approved) memiliki nilai feature importance tertinggi dibandingkan atribut lainnya. Selanjutnya, atribut Tuition fees up to date dan Curricular units 1st semester (approved) menempati urutan kedua dan ketiga sebagai faktor yang paling berkontribusi terhadap hasil klasifikasi. Selain ketiga atribut tersebut, fitur Curricular units 2nd semester (enrolled), Scholarship holder, Debtor, Curricular units 2nd semester (credited), Nationality, Gender, dan Course juga memberikan kontribusi terhadap proses prediksi, meskipun dengan tingkat kepentingan yang lebih rendah.

Hasil tersebut menunjukkan bahwa model XGBoost tidak hanya memanfaatkan informasi mengenai performa akademik mahasiswa, tetapi juga mempertimbangkan aspek finansial dan karakteristik mahasiswa dalam membedakan kelas Dropout dan Non-Dropout. Dengan demikian, feature importance memberikan gambaran mengenai atribut yang paling dominan digunakan model dalam proses klasifikasi dan menjadi dasar untuk analisis lebih lanjut pada bagian pembahasan[33].

3.3.1 Pembahasan

Kemampuan model dalam mengenali mahasiswa yang berisiko dropout dipengaruhi oleh proses penyeimbangan data menggunakan Hybrid SMOTE-ENN. Pembentukan sampel sintetis pada kelas minoritas melalui SMOTE diikuti dengan proses penyaringan data menggunakan ENN menghasilkan distribusi data latih yang lebih representatif. Kondisi tersebut mengurangi dominasi kelas mayoritas sehingga pola karakteristik mahasiswa dropout dapat dipelajari dengan lebih baik oleh model.

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

Berdasarkan hasil feature importance, atribut Curricular Units 2nd Semester (Approved) menempati posisi paling dominan dalam proses klasifikasi. Hal tersebut menunjukkan bahwa tingkat keberhasilan mahasiswa menyelesaikan mata kuliah pada semester kedua menjadi indikator penting dalam membedakan mahasiswa yang berpotensi dropout dan yang mampu melanjutkan studi. Capaian akademik pada semester awal dapat mencerminkan kemampuan adaptasi mahasiswa terhadap tuntutan pembelajaran di perguruan tinggi sehingga layak dijadikan indikator dalam sistem deteksi dini.

Kontribusi yang tinggi juga ditunjukkan oleh atribut Tuition Fees Up to Date, yang memperlihatkan bahwa aspek finansial memiliki hubungan dengan keberlangsungan studi mahasiswa. Kelancaran pembayaran biaya pendidikan dapat mencerminkan stabilitas mahasiswa dalam mengikuti proses akademik, sedangkan keterlambatan pembayaran berpotensi menjadi salah satu indikator meningkatnya risiko dropout. Oleh karena itu, identifikasi mahasiswa berisiko tidak cukup dilakukan berdasarkan indikator akademik saja, tetapi perlu mempertimbangkan kondisi finansial sebagai faktor pendukung.

Temuan tersebut memperlihatkan bahwa keberhasilan prediksi tidak hanya ditentukan oleh pemilihan algoritma klasifikasi, tetapi juga oleh kualitas data hasil proses resampling. Informasi mengenai faktor-faktor dominan tersebut dapat dimanfaatkan oleh perguruan tinggi sebagai dasar dalam menyusun strategi pendampingan akademik, pemberian bantuan finansial, serta pengembangan sistem deteksi dini untuk menekan angka dropout.

4. KESIMPULAN

Penerapan Hybrid Synthetic Minority Oversampling Technique–Edited Nearest Neighbour (SMOTE-ENN) yang dikombinasikan dengan algoritma Decision Tree, Random Forest, XGBoost, dan Voting Ensemble menunjukkan bahwa penanganan class imbalance berperan penting dalam meningkatkan kemampuan model dalam mengidentifikasi mahasiswa yang berisiko dropout. Hasil analisis feature importance mengindikasikan bahwa keberhasilan akademik pada semester awal serta kelancaran pembayaran biaya pendidikan merupakan faktor yang paling berkontribusi terhadap risiko kegagalan akademik. Temuan tersebut memperlihatkan bahwa pendekatan machine learning tidak hanya berfungsi sebagai metode klasifikasi, tetapi juga dapat dimanfaatkan sebagai dasar dalam pengembangan sistem deteksi dini dan penyusunan kebijakan akademik yang lebih tepat sasaran. Meskipun demikian, kajian ini masih terbatas pada penggunaan satu dataset publik dan belum mengintegrasikan pendekatan Explainable Artificial Intelligence (XAI) untuk meningkatkan interpretabilitas model. Oleh karena itu, penelitian selanjutnya disarankan memanfaatkan dataset dari berbagai perguruan tinggi, membandingkan teknik penanganan class imbalance lainnya, serta mengintegrasikan metode XAI, seperti SHAP atau LIME, agar interpretasi model menjadi lebih komprehensif dan dapat mendukung proses pengambilan keputusan akademik secara lebih efektif.

REFERENCES

[1] D. Anggraini, P. Korespondensi, and P. Mesin, “Data Mining Pendidikan: Predeksi Gaya Belajar Mahasiswa Teknik Menggunakan Machine Learning,” J. Teknol. Inf. dan Ilmu Komput., vol. 11, no. 3, pp. 563–572, 2025, doi: https://doi.org/10.25126/jtiik.2025129190.

[2] Z. Parinzka, S. Surono, and A. Thobirin, “Algortima Dupport Vector Regression dan Analisis Long Short-Term Memory Sebagai Penaganan Missing Data,” J. Teknol. Inf. dan Ilmu Komput., vol. 13, no. 1, pp. 73–82, 2026, doi: https://doi.org/10.25126/jtiik.2026131.

[3] M. A. Sadat, A. Pambudi, S. Ibad, U. D. Nuswantoro, I. Systems, and S. Program, “Comparison of Algoritm Between Classfication & Regression Trees and Support Vector Machine in Determining Student Accptance in State Universities,” J. Tek. Inform., vol. 4, no. 6, pp. 1589–1604, 2024, doi: https://doi.org/10.52436/1.jutif.2023.4.6.1565.

[4] N. Mubarak, S. Susandri, and A. Zamsuri, “Geodetically-Enhanced Hybrid GRU with Adaptive Dropout and Dynamic L2 Regularization for Earthquake Parameter Prediction in Indonesia,” J. Tek. Inform., vol. 7, no. 2, pp. 1483–1499, 2026, doi: https://doi.org/10.52436/1.jutif.2026.7.2.5478.

[5] N. Fadilah et al., “Perbandingan Kinerja Algortima Klasifikasi untuk Pemilihan Tempat Promosi Upaya Meningkatkan Jumlah Mahasiswa Baru,” J. Teknol. Inf. dan Ilmu Komput., vol. 13, no. 1, pp. 1–10, 2026, doi: https://doi.org/10.25126/jtiik.2026131.

[6] F. Maulana and E. B. Setiawan, “Performance of Deep Feed-Forward Neural Network Algorithm Based on Content-Based Filtering Approach,” INTENSIF, vol. 8, no. 2, pp. 278–294, 2024, doi: https://doi.org/10.29407/intensif.v8i2.22904.

[7] J. J. Purnama et al., “Klasifikasi Mahasiswa Her Berbasis Algoritma SVM dan Decision Tree,” J. Teknol. Inf. dan Ilmu Komput., vol. 7, no. 6, pp. 1253–1260, 2020, doi: 10.25126/jtiik.202073080.

[8] T. Suprapti, B. Nurhakim, B. Warni, A. Hermina, and V. A. Syahputra, “A Decision Tree Model with Grid Search Optimization for Scholarship Recipient Classification,” J. Tek. Inform., vol. 6, no. 5, pp. 3800–3813, 2025, doi: https://doi.org/10.52436/1.jutif.2025.6.5.5235.

[9] E. J. Kusuma, R. Nurmandhani, I. Pantiawati, and Y. M. Manglapy, “A Random Forest and SMOTE-Based Machine Learning Model for Predicting Recurrence in Papillary Thyroid Carcinoma,” J. Tek. Inform., vol. 6, no. 4, pp. 2019–2034, 2025, doi: https://doi.org/10.52436/1.jutif.2025.6.4.4854.

[10] M. Rizki, A. Hermawan, and D. Avianto, “Learning Accuracy with Particle Swarm Optimization for Music Genre Classification Using Recurrent Neural Networks,” Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 23, no. 2, pp. 297–308, 2024, doi: 10.30812/matrik.v23i2.3037.

[11] Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, “Dampak SMOTE terhadap Kinerja Random Forest Classifier

Journal of Information System Research (JOSH)
Volume 7, No. 3, April 2026, pp 798−807
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i3.9448

berdasarkan Data Tidak seimbang Impact of SMOTE on Random Forest Classifier Performance based on Imbalanced Data," Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 21, no. 3, 2022, doi: 10.30812/matrik.v21i3.1726.

[12] I. G. T. Suryawan, I. P. Agus, and E. Darma, "Optimasi Convulutional Neural Netwoek untuk Deteksi Covid-19 pada X-RAY Thorax Berbasis Dropout," J. Teknol. Inf. dan Ilmu Komput., vol. 9, no. 3, pp. 551–558, 2022, doi: 10.25126/jtiik.202295143.

[13] M. Lalu Ganda Rady Putra, Didik Dwi Prasetya, "Student Dropout Prediction Using Random Forest and XGBoost Method," INTENSIF, vol. 9, no. 1, pp. 147–157, 2025, doi: https://doi.org/10.29407/intensif.v9i1.21191.

[14] M. Ricky, P. Putra, and E. Utami, "Comparative Analysis of Hybrid Model Performance Using Stacking and Blending Techniques for Student Drop-Out Prediction in MOOC," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 158, pp. 1–6, 2026, doi: https://doi.org/10.29207/resti.v8i3.5760.

[15] S. H. Hasanah, E. Julianti, U. Terbuka, S. Informasi, and U. Terbuka, "Analysis of CART and Random Forest on Statistics Student Status at Universitas Terbuka," INTENSIF, vol. 6, no. 1, pp. 56–65, 2022, doi: https://doi.org/10.29407/intensif.v6i1.16156.

[16] A. A. Nababan, M. Jannah, M. Aulina, and D. Andrian, "Prediksi Kualitas Udara menggunakan XGBoost dengan Synthetic Minority Overampling Technique (SMOTE) Beradasarkan Indeks Standar Pencemaran Udara (ISPU)," J. Tek. Inform. Kaputama, vol. 7, no. 1, pp. 214–219, 2023, doi: https://doi.org/10.59697/jtik.v7i1.66.

[17] A. Anggrawan, H. Hairani, and C. Satria, "Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE," Int. J. Inf. Educ. Technol., vol. 13, no. 2, 2023, doi: 10.18178/ijiet.2023.13.2.1806.

[18] V. Realinho, M. Vieira Martins, J. Machado, and L. Baptista, "Predict Students Dropout and Academic Success," 2021, UCI Machine Learning Repository, Irvine, California. doi: 10.24432/C5MC89.

[19] T. Gori et al., "Preprocessing Data dan Klasifikasi untuk Prediksi Akademik Siswa," J. Teknol. Inf. dan Ilmu Komput., vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.

[20] S. Sidiq, N. S. Mabrur, S. T. Informatika, F. Teknik, and U. M. Tangerang, "Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost dan Teknik Oversampling," J. Ilm. Inform. dan Ilmu Komput., vol. 4, pp. 13–23, 2025, doi: https://doi.org/10.58602/jima-ilkom.v4i1.41.

[21] D. Kurniadi, F. Nuraeni, M. Firmansyah, K. Garut, and P. Korespondensi, "Klasifikasi Masyarakat Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Naive Bayes dan SMOTE," J. Teknol. Inf. dan Ilmu Komput., vol. 10, no. 2, pp. 309–320, 2023, doi: 10.25126/jtiik.2023106453.

[22] H. Ali, N. Hendrastuty, C. Science, and U. T. Indonesia, "Comparison of Naive Bayes Classdifier, Support Vector Machine, Rando Forest Algorithms Public Sentiment Analysis od KIP-K Program on Twitter," J. Tek. Inform., vol. 5, no. 6, pp. 1701–1712, 2024, doi: https://doi.org/10.52436/1.jutif.2024.5.6.4030.

[23] R. Siringoringo, D. Arisandi, E. Kurniawan, and E. B. Nababan, "Model Klasifikasi dengan Logistic Regression dan Recursive Feature Elimation pada Data Tidak Seimbang," J. Teknol. Inf. dan Ilmu Komput., vol. 11, no. 4, pp. 735–742, 2024, doi: 10.25126/jtiik.1148198.

[24] D. Kurnia et al., "Seleksi Fitur Dengan Particle Swarm Optimization pada Klasifikasi Penyakit Parkinson Menggunakan XBboost," J. Teknol. Inf. dan Ilmu Komput., vol. 10, no. 5, pp. 1083–1094, 2023, doi: 10.25126/jtiik.2023107252.

[25] M. D. Simbolon, D. D. Bukit, R. E. Purba, and F. H. Ketaren, "Perbandingan Algoritma K-Nearest Neighboor dan Naive Bayes Dalam Prediksi Penyakit Ginjal Kronis Pada Lansia," J. Inf. Syst. Res., vol. 6, no. 2, pp. 1519–1525, 2025, doi: 10.47065/josh.v6i2.5938.

[26] E. Junianto and S. Nurkhodijah, "Explainable Ensemble Learning for Depression Risk Classification Using Multidomain Behavioral Features," J. Tek. Inform., vol. 7, no. 2, pp. 778–792, 2026.

[27] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, "Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning Comparison of Feature Extraction in Multilabel Text Classification Using Machine Learning Algorithm," Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 21, no. 3, pp. 653–66, 2022, doi: 10.30812/matrik.v21i3.1851.

[28] W. S. Pratama, D. D. Prasetya, T. Widyaningtyas, M. Z. Wiryawan, and L. G. Rady, "Performance Evaluation of Artificial Intelligence Models for Classification in Concept Map Quality Assessment," Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 24, no. 3, pp. 407–422, 2025, doi: 10.30812/matrik.v24i3.4729.

[29] A. H. M. Suparyati, Emma Utami, "Applying Different Resampling Strategies In Random Forest Algorithm To Predict Lumpy Skin Disease," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 158, pp. 555–562, 2026, doi: https://doi.org/10.29207/resti.v6i4.4147.

[30] I. Raga et al., "Analisis Sentimen Pasien Terhadap Layanan Anatrmedika DentalCare Menggunakan Metode XGBoost," J. Teknol. Inf. dan Ilmu Komput., vol. 12, no. 6, pp. 1377–1384, 2025, doi: https://doi.org/10.25126/jtiik.2025126.

[31] N. Putri et al., "Penerapan Feature Enngineering dan Hyperparameter Tuning untuk Meningkatkan Akurasi Model Random Forest Pada Klasifikasi Risiko Kredit," J. Teknol. Inf. dan Ilmu Komput., vol. 12, no. 2, pp. 251–262, 2025, doi: 10.25126/jtiik.2025128472.

[32] D. Kurniadi, A. I. Zulkarnaen, A. Mulyani, K. Garut, and P. Korespondensi, "Perancangan Model Convolutional Neural Network pada Aplikasi Pengenalan Askara Sunda Berbasis Mobile," J. Teknol. Inf. dan Ilmu Komput., vol. 12, no. 5, 2025, doi: https://doi.org/10.25126/jtiik.2025125.

[33] D. W. W. Haryono Setiadi , Indah Paksi Larasati, Esti Suryani, A. W. Hasan Dwi Cahyono, and A. Doewes, "Comparing Correlation-Based Feature Selection and Symmetrical," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 158, pp. 542–554, 2026, doi: https://doi.org/10.29207/resti.v8i4.5911.