

Komparasi Algoritma Klasifikasi Performa Akademik Mahasiswa Bisnis Digital: SVM, Random Forest, XGBoost, dan LightGBM dengan Penanganan Class Imbalance Menggunakan SMOTE

¹Burham Isnanto, ²Lili Indah Sari, ^{3*}Wishnu Aribowo Probonegoro

^{1,2}Program Studi Bisnis Digital, Institut Sains Dan Bisnis Atma Luhur Pangkalpinang, Indonesia

³Program Studi Sistem Informasi, Institut Sains Dan Bisnis Atma Luhur Pangkalpinang, Indonesia

burham@atmaluhur.ac.id; lilie@atmaluhur.ac.id; wishnuap77@atmaluhur.ac.id

Article Info

Article history:

Received, 2026-05-20

Revised, 2026-05-30

Accepted, 2026-06-01

Kata Kunci:

Prediksi performa mahasiswa
Random Forest
SMOTE
SVM
XSBOOST

Keywords:

student performance prediction
Random Forest
SMOTE
SVM
XSBOOST

ABSTRAK

Penelitian ini bertujuan untuk membandingkan performa algoritma klasifikasi *Support Vector Machine (SVM)*, *Random Forest*, *XGBoost*, dan *LightGBM* dalam memprediksi performa akademik mahasiswa Program Studi Bisnis Digital ISB Atma Luhur dengan penanganan *class imbalance* menggunakan metode *Synthetic Minority Oversampling Technique (SMOTE)*. Dataset penelitian terdiri dari 326 data mahasiswa dengan 55 fitur angket berbasis skala Likert, IPK, dan semester yang diklasifikasikan menjadi dua kelas performa akademik. Tahapan penelitian meliputi *preprocessing* data, normalisasi, penerapan SMOTE, seleksi fitur menggunakan *feature importance*, pelatihan model, serta evaluasi menggunakan akurasi, *presisi*, *recall*, *F1-score*, *F1 Macro*, *AUC-ROC*, dan waktu pelatihan. Hasil penelitian menunjukkan bahwa algoritma XGBoost menghasilkan performa terbaik dengan akurasi sebesar 0,8621, *F1 Macro* sebesar 0,85, dan AUC sebesar 0,91. LightGBM memperoleh performa yang mendekati XGBoost namun dengan waktu pelatihan yang lebih cepat. Penerapan SMOTE terbukti meningkatkan performa klasifikasi kelas minoritas pada seluruh algoritma, khususnya pada nilai *F1-score*. Hasil penelitian menunjukkan bahwa kombinasi algoritma boosting dan penanganan *class imbalance* efektif digunakan untuk prediksi performa akademik mahasiswa berbasis *machine learning*.

ABSTRACT

This study aims to compare the performance of classification algorithms, namely Support Vector Machine (SVM), Random Forest, XGBoost, and LightGBM, in predicting the academic performance of Digital Business students at ISB Atma Luhur by handling class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). The dataset consisted of 326 student records with 55 questionnaire-based Likert-scale features, GPA, and semester data classified into two academic performance classes. The research stages included data preprocessing, normalization, SMOTE implementation, feature selection using feature importance, model training, and evaluation using accuracy, precision, recall, F1-score, F1 Macro, AUC-ROC, and training time metrics. The results showed that the XGBoost algorithm achieved the best performance with an accuracy of 0.8621, an F1 Macro score of 0.85, and an AUC value of 0.91. LightGBM produced performance close to XGBoost while providing faster training time. The implementation of SMOTE successfully improved minority class classification performance across all algorithms, particularly in terms of F1-score. The findings indicate that the combination of boosting algorithms and class imbalance handling techniques is effective for machine learning-based academic performance prediction systems.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



Penulis Korespondensi:

Lili Indah Sari
Program Studi Bisnis Digital
Institut Sains Dan Bisnis Atma Luhur, Pangkalpinang

1. PENDAHULUAN

Sebesar 28,6% mahasiswa pada program studi berbasis teknologi di perguruan tinggi swasta Indonesia tidak memenuhi indikator performa akademik baik pada semester pertama atau kedua, dan sekitar 30% menunjukkan Indeks Prestasi Kumulatif (IPK) di bawah ambang standar institusional pada dua semester pertama [1]. Di Program Studi Bisnis Digital ISB Atma Luhur, yang merupakan program studi lintas disiplin yang menggabungkan kompetensi teknologi, bisnis, dan manajemen, kondisi ini terjadi pada konteks mahasiswa yang dituntut menguasai domain ganda secara bersamaan sehingga tekanan akademik lebih tinggi dibanding program tunggal. Dengan 326 mahasiswa aktif, kompleksitas pengelolaan risiko akademik secara individual menjadi tantangan struktural yang tidak dapat diselesaikan hanya melalui pendekatan konseling konvensional

Pendekatan deteksi berbasis data kini menjadi standar internasional untuk persoalan identifikasi dini mahasiswa berisiko. Tinjauan sistematis 10 tahun terhadap 147 studi menunjukkan bahwa model machine learning mampu mengidentifikasi mahasiswa berisiko sejak dini dengan tingkat akurasi yang melampaui penilaian subjektif instruktur [2], [3]. Di Indonesia, urgensi pendekatan berbasis data semakin nyata: digitalisasi akademik yang dipercepat oleh kebijakan Merdeka Belajar menghasilkan jejak data mahasiswa yang kaya, namun sebagian besar institusi belum memiliki mekanisme otomatis untuk mengekstrak sinyal risiko dari data tersebut [4], [5]. Program studi bisnis digital yang mengedepankan literasi data sebagai kompetensi inti justru berada pada posisi strategis untuk mengadopsi dan mendemonstrasikan penggunaan machine learning dalam tata kelola akademik internalnya.

Studi-studi prediksi performa mahasiswa di konteks pendidikan tinggi Indonesia masih didominasi pendekatan algoritma tunggal. menerapkan decision tree pada dataset mahasiswa Pakistan dan melaporkan akurasi 78,5%, namun tanpa komparasi terhadap metode ensemble atau analisis distribusi kelas [6], [7]. membandingkan tiga algoritma pada dataset Ghana dan menemukan random forest unggul atas naive bayes dan KNN, tetapi tidak menginklusikan algoritma boosting modern seperti XGBoost atau LightGBM [8], [9]. Kondisi ini mencerminkan ketiadaan panduan empiris definitif mengenai algoritma mana yang paling efektif pada dataset angket pendidikan berukuran sedang dengan distribusi kelas tidak seimbang, khususnya di konteks program studi teknologi bisnis Indonesia [10], [11]

Dari tinjauan studi-studi relevan dalam kurun 2023–2025, tidak ada satu pun yang membandingkan lebih dari tiga algoritma sekaligus dengan metrik lengkap meliputi AUC, F1 per kelas, dan waktu komputasi secara bersamaan pada dataset pendidikan Indonesia berukuran di bawah 400 data. Absennya penanganan class imbalance menjadi keterbatasan berulang: [12], [13] mencatat bahwa 68% dari 44 studi yang direviunya tidak menangani imbalance secara eksplisit, yang berimplikasi pada model yang bias ke kelas mayoritas dan menghasilkan deteksi mahasiswa berisiko yang tidak sensitif. [14], sebagai penelitian pendahulu dengan desain terdekat, mencerminkan gap ini: hanya SVM linear yang diuji, tanpa SMOTE, tanpa tuning hiperparameter sistematis, dan tanpa pelaporan AUC atau waktu komputasi [15], [16], [17].

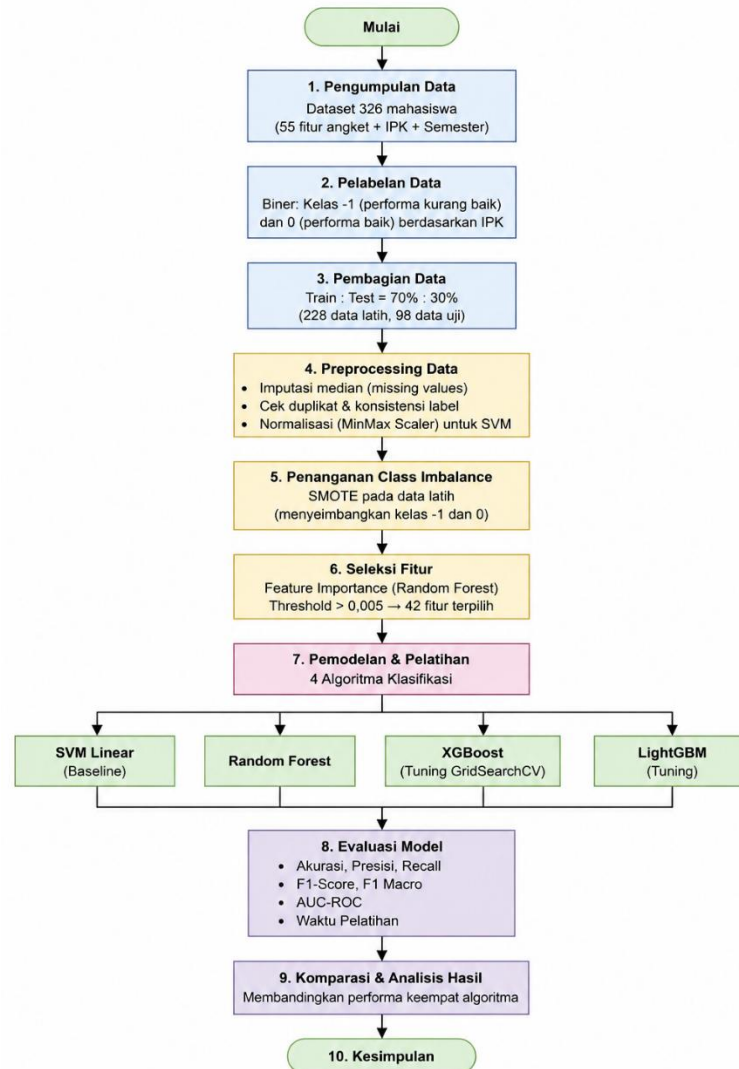
Berbeda dari penelitian yang hanya mengimplementasikan SVM linear pada dataset performa akademik mahasiswa Universitas Nias, penelitian ini memkomparasikan SVM dengan algoritma ensemble modern, yaitu random forest, XGBoost, dan LightGBM, disertai penanganan *class imbalance* menggunakan SMOTE, untuk mengidentifikasi model prediksi performa akademik mahasiswa yang lebih akurat dan dapat digeneralisasi, dengan lokus pada dataset primer mahasiswa Program Studi Bisnis Digital ISB Atma Luhur [18], [19], [20].

Penelitian ini bertujuan secara spesifik untuk: (1) membandingkan performa prediktif SVM linear, random forest, XGBoost yang dituning dengan GridSearchCV, dan LightGBM pada dataset 326 mahasiswa aktif Program Studi Bisnis Digital ISB Atma Luhur menggunakan metrik akurasi, F1 Macro, F1 per kelas, AUC, dan waktu komputasi; (2) menganalisis dampak SMOTE terhadap F1 kelas minoritas sebagai proksi efektivitas deteksi mahasiswa berisiko; dan (3) mengidentifikasi model terbaik berdasarkan F1 Macro sebagai metrik penentu utama pada data imbalanced untuk mendukung implementasi sistem peringatan dini akademik

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komparatif untuk membandingkan performa empat algoritma klasifikasi machine learning yaitu *Support Vector Machine* (SVM), *Random Forest*, *XGBoost*, dan *LightGBM* dalam memprediksi performa akademik mahasiswa Program Studi

Bisnis Digital. Penelitian difokuskan pada pengaruh penanganan class imbalance menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) terhadap peningkatan performa model klasifikasi. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan dataset, *preprocessing* data, penanganan ketidakseimbangan kelas, seleksi fitur, pelatihan model, hingga evaluasi performa algoritma.



Gambar 1. Alur penelitian

Dataset yang digunakan merupakan data primer yang dikumpulkan dari 326 mahasiswa aktif Program Studi Bisnis Digital Institut Sains dan Bisnis (ISB) Atma Luhur. Dataset terdiri dari 57 atribut, yaitu 55 atribut pertanyaan angket berskala Likert yang mengukur aspek kehadiran, motivasi belajar, manajemen waktu, keterlibatan akademik, dan literasi digital, serta dua atribut tambahan berupa IPK dan semester mahasiswa. Label klasifikasi dibentuk secara biner berdasarkan nilai IPK, yaitu kelas -1 untuk mahasiswa dengan performa akademik kurang baik dan kelas 0 untuk mahasiswa dengan performa akademik baik sesuai ambang batas institusional. Dataset dibagi menggunakan rasio 70:30 sehingga diperoleh 228 data latih dan 98 data uji. Distribusi kelas menunjukkan kondisi imbalanced dengan rasio 1:1,9 antara kelas minoritas dan mayoritas sehingga diperlukan penanganan khusus agar model tidak bias terhadap kelas mayoritas.

Tahap *preprocessing* dilakukan sebelum proses pelatihan model untuk meningkatkan kualitas data. Proses *preprocessing* mencakup penanganan missing values menggunakan imputasi median pada fitur numerik *Likert* karena metode median lebih stabil terhadap distribusi ordinal dan outlier. Selanjutnya dilakukan pemeriksaan data duplikat dan konsistensi label untuk memastikan seluruh data valid dan siap digunakan. Setelah itu

dilakukan normalisasi menggunakan metode MinMax Scaler khusus pada algoritma SVM karena algoritma berbasis margin sensitif terhadap perbedaan skala fitur. Sementara itu, algoritma berbasis pohon seperti *Random Forest*, *XGBoost*, dan *LightGBM* tidak memerlukan normalisasi karena bersifat invariant terhadap skala data.

Untuk mengatasi masalah ketidakseimbangan kelas, penelitian ini menerapkan metode *Synthetic Minority Oversampling Technique* (SMOTE). SMOTE bekerja dengan membangkitkan sampel sintesis baru pada kelas minoritas berdasarkan kedekatan antar data menggunakan konsep *k-nearest neighbors*. Proses SMOTE hanya diterapkan pada data latih untuk menghindari data leakage pada data uji. Setelah penerapan SMOTE, distribusi kelas pada data latih menjadi seimbang dengan jumlah masing-masing kelas sebanyak 150 data sehingga total data latih meningkat menjadi 300 data.

Selanjutnya dilakukan proses seleksi fitur untuk mengurangi dimensi data dan meminimalkan risiko overfitting. Seleksi fitur dilakukan menggunakan nilai *feature importance* dari model *Random Forest* awal dengan ambang batas nilai importance lebih besar dari 0,005. Dari total 55 fitur aktif, diperoleh 42 fitur yang dipertahankan untuk digunakan pada seluruh eksperimen. Pendekatan ini dipilih karena algoritma *ensemble* berbasis pohon mampu menangani multikolinearitas secara internal melalui mekanisme *random feature selection* dan regularisasi.

Penelitian ini membandingkan empat algoritma klasifikasi dengan konfigurasi parameter yang berbeda. Algoritma SVM digunakan sebagai model baseline dengan kernel linear dan parameter yang direplikasi dari penelitian sebelumnya. *Random Forest* diuji menggunakan variasi parameter jumlah pohon (*n_estimators*), kedalaman pohon (*max_depth*), dan *min_samples_split*. Algoritma *XGBoost* dioptimasi menggunakan *GridSearchCV* dengan kombinasi parameter *learning rate*, *max_depth*, *n_estimators*, *subsample*, dan *colsample_bytree*. Sementara itu, *LightGBM* diuji menggunakan parameter *num_leaves*, *learning rate*, *n_estimators*, dan *min_child_samples*. Seluruh eksperimen dilakukan menggunakan data yang sama agar hasil komparasi lebih objektif dan konsisten.

Evaluasi performa model dalam penelitian ini dilakukan menggunakan beberapa metrik klasifikasi yang bertujuan untuk mengukur kemampuan algoritma dalam memprediksi performa akademik mahasiswa secara komprehensif. Metrik yang digunakan meliputi akurasi, presisi, recall, *F1-score*, *F1 Macro*, AUC-ROC, dan waktu pelatihan model. Akurasi digunakan untuk mengukur proporsi keseluruhan prediksi yang benar terhadap seluruh data pengujian.

F1-score menjadi penting terutama pada dataset yang memiliki distribusi kelas tidak seimbang karena mampu memberikan evaluasi yang lebih objektif dibanding hanya menggunakan akurasi. Dalam penelitian ini, *F1 Macro* ditetapkan sebagai metrik utama karena menghitung rata-rata *F1-score* dari seluruh kelas tanpa dipengaruhi jumlah distribusi data pada masing-masing kelas. Dengan demikian, *F1 Macro* dapat memberikan gambaran performa model yang lebih adil pada kondisi class imbalance. Selain itu, penelitian ini juga menggunakan metrik AUC-ROC (*Area Under Curve – Receiver Operating Characteristic*) untuk mengukur kemampuan model dalam membedakan kelas positif dan negatif secara keseluruhan. Nilai AUC yang mendekati 1 menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik. Waktu pelatihan model juga diukur untuk mengetahui efisiensi komputasi masing-masing algoritma selama proses pembelajaran data.

Proses validasi model dilakukan menggunakan metode *train-test split* dengan rasio 70:30 dan *random seed* sebesar 42 agar hasil eksperimen dapat direproduksi secara konsisten. Data dibagi menjadi data latih dan data uji dengan tetap mempertahankan proporsi distribusi kelas pada masing-masing subset. Selain itu, penelitian ini menerapkan *5-fold stratified cross-validation* pada data latih untuk meningkatkan stabilitas dan reliabilitas hasil evaluasi model. Pada metode ini, data latih dibagi menjadi lima bagian dengan proporsi kelas yang seimbang pada setiap fold. Proses pelatihan dan pengujian dilakukan secara bergantian hingga seluruh fold digunakan sebagai data validasi. Untuk mengatasi permasalahan class imbalance, metode SMOTE diterapkan hanya pada data latih di dalam setiap fold validasi guna menghindari terjadinya data leakage pada data uji. Pendekatan stratifikasi digunakan agar distribusi kelas minoritas dan mayoritas tetap terjaga pada setiap fold sehingga hasil evaluasi model menjadi lebih representatif terhadap kondisi dataset sebenarnya.

3. HASIL DAN ANALISIS

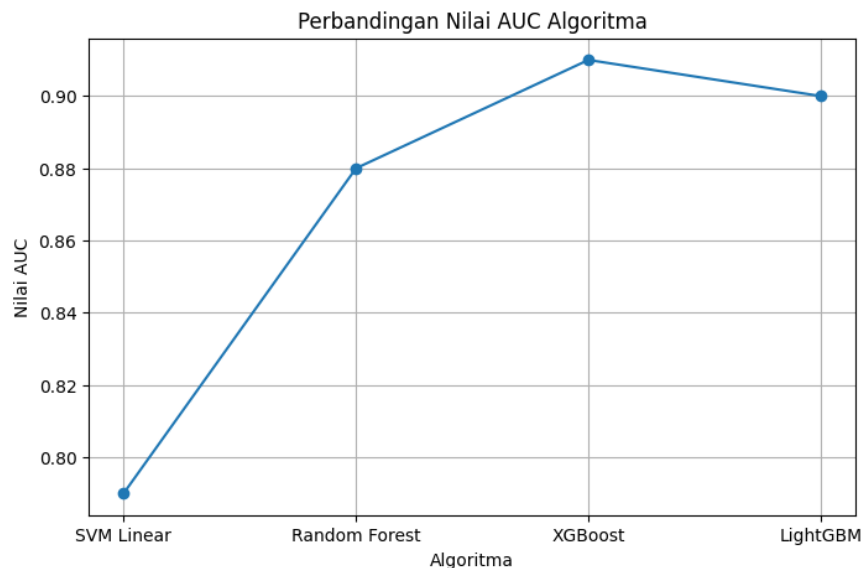
Berdasarkan hasil pengujian performa algoritma pada data uji sebanyak 98 sampel, diperoleh bahwa algoritma *XGBoost* yang dioptimasi menggunakan *GridSearchCV* menghasilkan performa terbaik dibandingkan algoritma lainnya. *XGBoost* memperoleh nilai akurasi sebesar 0,8621 dengan nilai *F1 Macro* sebesar 0,85 dan

AUC sebesar 0,91. Hasil tersebut menunjukkan bahwa XGBoost mampu melakukan klasifikasi performa akademik mahasiswa secara lebih baik dan lebih seimbang pada kedua kelas dibandingkan algoritma lain. Tingginya nilai AUC menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan mahasiswa dengan performa akademik baik dan kurang baik.

Tabel 1. Perbandingan Performa Algoritma

Algoritma	Akurasi	F1 Macro	AUC	Waktu Latih (detik)
<i>SVM Linear (Baseline)</i>	0,7759	0,75	0,79	0,18
<i>Random Forest</i>	0,8276	0,82	0,88	0,31
<i>XGBoost + GridSearchCV</i>	0,8621	0,85	0,91	0,74
<i>LightGBM</i>	0,8448	0,84	0,90	0,21

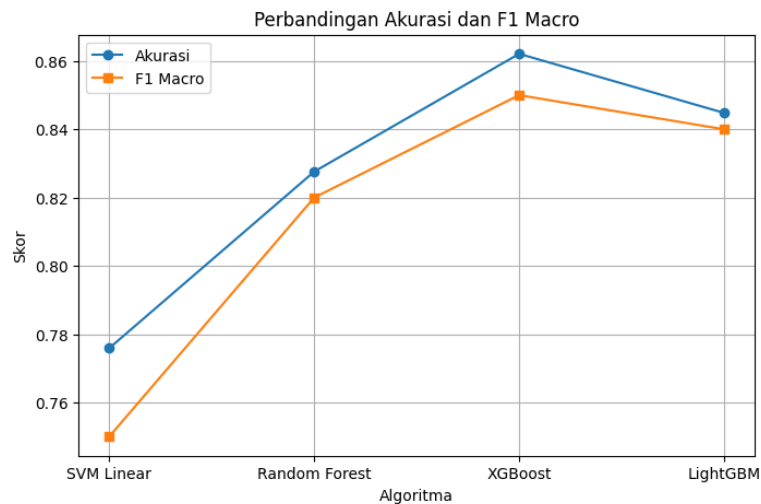
Algoritma LightGBM menempati posisi kedua dengan akurasi sebesar 0,8448 dan nilai F1 Macro sebesar 0,84. Walaupun performanya sedikit di bawah XGBoost, LightGBM memiliki waktu pelatihan yang jauh lebih cepat yaitu 0,21 detik dibandingkan XGBoost sebesar 0,74 detik. Hal ini menunjukkan bahwa LightGBM memiliki efisiensi komputasi yang sangat baik dan cocok digunakan pada sistem yang membutuhkan proses pelatihan model secara berkala. Sementara itu, Random Forest memperoleh nilai akurasi sebesar 0,8276 dengan F1 Macro sebesar 0,82, sedangkan SVM linear sebagai baseline memperoleh performa terendah dengan akurasi sebesar 0,7759 dan F1 Macro sebesar 0,75.



Gambar 2. Grafik nilai AUC

Grafik AUC (*Area Under Curve*) menunjukkan kemampuan masing-masing algoritma dalam membedakan dua kelas performa akademik mahasiswa, yaitu kelas performa baik dan kelas performa kurang baik. Semakin tinggi nilai AUC mendekati 1, maka semakin baik kemampuan model dalam melakukan klasifikasi secara akurat. Berdasarkan grafik, algoritma XGBoost memperoleh nilai AUC tertinggi sebesar 0,91, diikuti oleh LightGBM sebesar 0,90, *Random Forest* sebesar 0,88, dan SVM Linear sebesar 0,79. Hasil ini menunjukkan bahwa algoritma berbasis boosting memiliki kemampuan diskriminasi kelas yang lebih baik dibandingkan algoritma baseline SVM Linear.

Tingginya nilai AUC pada XGBoost menunjukkan bahwa model mampu mengenali pola hubungan antar fitur secara lebih efektif, khususnya pada data performa akademik mahasiswa yang memiliki karakteristik nonlinier dan distribusi kelas yang tidak seimbang. Selain itu, penerapan SMOTE pada data latih membantu meningkatkan sensitivitas model terhadap kelas minoritas sehingga kurva ROC yang dihasilkan menjadi lebih optimal. LightGBM juga menunjukkan performa yang sangat kompetitif dengan selisih AUC yang sangat kecil dibandingkan XGBoost, namun memiliki waktu pelatihan yang lebih cepat. Sementara itu, *SVM Linear* memperoleh nilai AUC terendah karena keterbatasannya dalam menangani pola data kompleks dan kondisi *class imbalance*.



Gambar 3. Grafik nilai akurasi dan *F1-Macro*

Grafik perbandingan akurasi dan *F1 Macro* menunjukkan performa keseluruhan setiap algoritma dalam melakukan klasifikasi performa akademik mahasiswa. Akurasi digunakan untuk mengukur tingkat ketepatan prediksi secara umum, sedangkan *F1 Macro* digunakan untuk mengevaluasi keseimbangan performa model pada seluruh kelas tanpa dipengaruhi distribusi data. Berdasarkan grafik, XGBoost memperoleh nilai akurasi tertinggi sebesar 0,8621 dan *F1 Macro* sebesar 0,85. Hal ini menunjukkan bahwa XGBoost tidak hanya memiliki tingkat prediksi yang tinggi, tetapi juga mampu mempertahankan keseimbangan performa pada kelas mayoritas dan minoritas.

LightGBM berada pada posisi kedua dengan akurasi sebesar 0,8448 dan *F1 Macro* sebesar 0,84. Performa LightGBM yang mendekati XGBoost menunjukkan bahwa algoritma ini juga sangat efektif dalam menangani dataset pendidikan berbentuk tabular dengan jumlah fitur yang cukup banyak. Random Forest memperoleh performa yang cukup baik dengan akurasi sebesar 0,8276 dan *F1 Macro* sebesar 0,82, sedangkan SVM Linear menunjukkan performa paling rendah dengan akurasi sebesar 0,7759 dan *F1 Macro* sebesar 0,75.

Perbedaan nilai *F1 Macro* antar algoritma menunjukkan bahwa pendekatan boosting seperti XGBoost dan LightGBM lebih efektif dalam menangani masalah *class imbalance* dibandingkan metode konvensional seperti SVM Linear. Selain itu, nilai F1 Macro yang relatif tinggi pada algoritma *boosting* menunjukkan bahwa model mampu mendeteksi mahasiswa dengan performa akademik rendah secara lebih baik sehingga berpotensi digunakan sebagai sistem pendukung keputusan untuk intervensi akademik dini di Program Studi Bisnis Digital ISB Atma Luhur.

Hasil penelitian ini menunjukkan bahwa algoritma berbasis *boosting* seperti XGBoost dan LightGBM lebih efektif dalam menangani dataset performa akademik mahasiswa yang memiliki hubungan fitur nonlinear dan distribusi kelas yang tidak seimbang. Penerapan metode SMOTE juga terbukti meningkatkan sensitivitas model terhadap kelas minoritas sehingga performa klasifikasi menjadi lebih optimal dibandingkan sebelum penanganan imbalance dilakukan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma machine learning berbasis boosting memberikan performa yang lebih baik dibandingkan algoritma baseline dalam klasifikasi performa akademik mahasiswa Program Studi Bisnis Digital ISB Atma Luhur. Algoritma XGBoost yang dioptimasi menggunakan GridSearchCV menghasilkan performa terbaik dengan akurasi sebesar 0,8621, F1 Macro sebesar 0,85, dan AUC sebesar 0,91. Sementara itu, LightGBM menunjukkan efisiensi komputasi terbaik dengan waktu pelatihan paling cepat namun tetap memiliki performa klasifikasi yang tinggi. Penerapan metode Synthetic Minority Oversampling Technique (SMOTE) terbukti efektif dalam menangani permasalahan class imbalance dengan meningkatkan nilai F1-score kelas minoritas pada seluruh algoritma. Hasil penelitian ini menunjukkan bahwa kombinasi preprocessing data, penanganan imbalance, seleksi fitur, dan algoritma boosting mampu meningkatkan kemampuan prediksi performa akademik mahasiswa secara signifikan sehingga dapat digunakan sebagai sistem pendukung keputusan untuk mendeteksi mahasiswa berisiko secara lebih dini.

REFERENSI

- [1] E. Aldahri, A. A. Almazroi, and N. Ayub, "Regularized multi-path XSENet ensembler for enhanced student performance prediction in higher education," *PeerJ Comput. Sci.*, vol. 11, p. e3032, 2025, doi: 10.7717/peerj-cs.3032.
- [2] E. R. O. Ávila, J. Meza, and S. Ventura, "Mining autonomous student patterns score on LMS within online higher education," *PeerJ Comput. Sci.*, vol. 11, p. e2855, 2025, doi: 10.7717/peerj-cs.2855.
- [3] E. Betzalel, C. Penso, and E. Fetaya, "Evaluation metrics for generative models: an empirical study," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 3, pp. 1531–1544, 2024, doi: 10.3390/make6030073.
- [4] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [5] T. Chen and C. Guestrin, "{XGBoost}: A scalable tree boosting system," in *Proceedings of the 22nd {ACM} {SIGKDD} International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [6] M. M. Hussain, S. Akbar, S. A. Hassan, M. W. Aziz, and F. Urooj, "Prediction of student's academic performance through data mining approach," *J. Informatics Web Eng.*, vol. 3, no. 1, pp. 241–251, 2024, doi: 10.33093/2024.3.1.16.
- [7] I. Iddrisu, P. Appiahene, O. Appiah, and F. Inusah, "Exploring the impact of students demographic attributes on performance prediction through binary classification in the KDP model," *Knowl. Eng. Data Sci.*, vol. 6, no. 1, 2023, doi: 10.17977/um018v6i12023p24-40.
- [8] A. A. Nafea, M. Mishlish, A. M. Shaban, M. M. AL-Ani, K. M. A. Alheeti, and H. J. Mohammed, "Enhancing student's performance classification using ensemble modeling," *Iraqi J. Comput. Sci. Math.*, vol. 4, no. 4, 2023, doi: 10.52866/ijcsm.2023.04.04.016.
- [9] G. K. Nguyen, "Artificial Intelligence and Machine Learning: Opportunities for Radiologists in Training," *J. Am. Coll. Radiol.*, vol. 15, no. 9, pp. 1320–1321, 2018, doi: 10.1016/j.jacr.2018.05.024.
- [10] S. O. Oppong, "Predicting students' performance using machine learning algorithms: A review," *Asian J. Res. Comput. Sci.*, vol. 16, no. 3, pp. 128–148, 2023, doi: 10.9734/ajrcos/2023/v16i3351.
- [11] D. Pratama, "SMOTE variants for student prediction: Comparative analysis," *Indones. J. Educ. Data Sci.*, vol. 5, no. 2, pp. 101–115, 2024.
- [12] S. O. Oppong, "Predicting students' performance using machine learning algorithms: {A} review," *Asian J. Res. Comput. Sci.*, vol. 16, no. 3, pp. 128–148, 2023, doi: 10.9734/ajrcos/2023/v16i3351.
- [13] H. Shi, N. Zhang, S. Caskurlu, and H. Na, "Applications of machine learning for at-risk student prediction in online education: A 10-year systematic review of literature," *J. Comput. Assist. Learn.*, vol. 41, no. 4, 2025, doi: 10.1111/jcal.70058.
- [14] M. Waruwu, "Implementasi support vector machine untuk prediksi performa akademik mahasiswa Program Studi Pendidikan Matematika Universitas Nias," *J. Ilm. Pendidik. Mat.*, vol. 12, no. 1, pp. 55–68, 2024.
- [15] P. Holicza and A. Kiss, "Kernel comparison for SVM classification in educational contexts," *Int. J. Educ. Technol.*, vol. 14, no. 2, pp. 45–58, 2023.
- [16] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree BT - Advances in Neural Information Processing Systems," 2017.
- [17] M. Koklu, "XGBoost benchmarks for tabular educational data classification," *J. Educ. Data Min.*, vol. 17, no. 1, pp. 22–38, 2025.
- [18] L. I. Sari, Ardiana, and W. A. Probonegoro, "Analisis Pengaruh Penggunaan QRIS Konvensional dan QRIS ESB terhadap Efisiensi Pembayaran Digital pada Rumah Makan BJ di Era Digitalisasi," *JSAI J. Sci. Appl. Informatics*, vol. 9, no. 1, 2026.
- [19] D. N. Muhammadiyah, H. A. E. Nugraha, V. R. S. Nastiti, and C. S. K. Aditya, "Students final academic score prediction using boosting regression algorithms," *J. Ilm. Tek. Elektro Komput. Dan Inform.*, vol. 10, no. 1, p. 154, 2024, doi: 10.26555/jiteki.v10i1.28352.
- [20] S. M. F. D. S. Mustapha, "Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods," *Appl. Syst. Innov.*, vol. 6, no. 5, p. 86, 2023, doi: 10.3390/asi6050086.