

Prediksi Kelulusan Siswa Sekolah Menengah Menggunakan XGBoost dan Analisis Feature Importance

¹*Luthfia Nurma Hapsari, ²Dziky Ridhwanullah, ³Miftakhurrokhmat

¹*Program Studi Sistem Informasi, Universitas Tiga Serangkai, Surakarta, Indonesia

²Program Studi Informatika, Universitas Tiga Serangkai, Surakarta, Indonesia

³Program Studi Teknologi Informasi, STMIK AMIKOM Surakarta, Surakarta, Indonesia

*Korespondensi: luthfianh@tsu.ac.id

Submit : 04 Jan 2026 | Diterima : 18 Mei 2026 | Terbit : 23 Mei 2026

ABSTRACT

Predicting the academic performance of secondary school students is an essential aspect of educational data mining, supporting decision-making in education and enabling early interventions for students at risk of failure. While machine learning approaches have been widely applied to student graduation prediction, most studies emphasize model accuracy and pay less attention to the interpretability of prediction results. In practice, understanding the factors influencing graduation is crucial for educators. This study aims to develop a prediction model for secondary school student graduation using the XGBoost algorithm and to identify the most influential features through feature importance analysis as a basis for educational intervention recommendations. The research was conducted within the framework of educational data mining, utilizing the Student Performance dataset from the UCI Machine Learning Repository, which includes 395 students and covers academic, social, and demographic variables. Data preprocessing involved categorical feature encoding and stratified data splitting with an 80:20 ratio. The XGBoost model was trained and evaluated using accuracy, precision, recall, and F1-score metrics, with particular attention to the minority class of at-risk students. Evaluation results show that the model achieved an accuracy of 0.71 and a Macro F1-score of 0.63. Feature importance analysis identified previous failures, guardian status, additional educational support, and absenteeism as the most influential factors in predicting student graduation. The findings indicate that the XGBoost model provides adequate predictive performance in analyzing student outcomes while also offering informative feature interpretation. The integration of educational data mining with feature importance analysis makes this model applicable as a decision-support tool and a foundation for designing data-driven educational interventions.

Keywords: Educational Data Mining; Feature Importance; Student Graduation Prediction; Student Performance; XGBoost

ABSTRAK

Prediksi performa akademik (student performance) siswa sekolah menengah merupakan bagian penting dalam kajian educational data mining untuk mendukung pengambilan keputusan pendidikan dan perancangan intervensi dini bagi siswa berisiko gagal. Pendekatan machine learning telah banyak digunakan untuk prediksi kelulusan siswa, namun sebagian besar penelitian masih berfokus pada akurasi model dan kurang menekankan aspek interpretabilitas hasil prediksi. Padahal, pemahaman faktor-faktor yang memengaruhi kelulusan sangat dibutuhkan oleh praktisi pendidikan. Penelitian ini bertujuan membangun model prediksi kelulusan siswa sekolah menengah menggunakan algoritma XGBoost serta mengidentifikasi fitur-fitur paling berpengaruh melalui analisis feature importance sebagai dasar rekomendasi intervensi pendidikan. Metode penelitian ini berada dalam konteks educational data mining dengan memanfaatkan dataset Student Performance dari UCI Machine Learning Repository yang terdiri dari 395 siswa dan mencakup variabel akademik, sosial, serta demografis. Data diproses melalui encoding fitur kategorikal dan pembagian data secara stratified dengan rasio 80:20. Model XGBoost dilatih dan dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score, dengan perhatian khusus pada kelas minoritas siswa berisiko gagal. Hasil evaluasi menunjukkan bahwa model mencapai accuracy sebesar 0,71 dan Macro F1-score sebesar 0,63. Analisis feature importance mengidentifikasi jumlah kegagalan sebelumnya, status wali,

dukungan pendidikan tambahan, dan jumlah ketidakhadiran sebagai faktor paling berpengaruh terhadap prediksi kelulusan siswa. Hasil penelitian menunjukkan bahwa model XGBoost mampu memberikan prediksi kelulusan siswa yang memadai dalam analisis student performance, sekaligus menyediakan interpretasi fitur yang informatif. Integrasi pendekatan educational data mining dengan analisis feature importance menjadikan model ini aplikatif sebagai alat bantu pengambilan keputusan dan perancangan intervensi pendidikan berbasis data.

Kata kunci: Educational Data Mining; Feature Importance; Prediksi Kelulusan Siswa; Student Performance; XGBoost

PENDAHULUAN

Performa akademik siswa merupakan indikator penting untuk menilai keberhasilan pendidikan dan merencanakan strategi intervensi yang tepat (Suleiman, Okunade, Dada, & Ezeanya, 2024). Evaluasi yang akurat terhadap performa ini membantu guru dan institusi pendidikan untuk mengidentifikasi siswa yang berisiko gagal serta merancang dukungan yang sesuai, seperti bimbingan tambahan atau program remedial (Ravindra, Yogita, Shrikrishna, Lalita, & Aniket, 2024). Tantangan dalam prediksi performa akademik mencakup heterogenitas kemampuan siswa, pengaruh faktor sosial dan keluarga, serta distribusi kelas yang tidak seimbang antara siswa yang lulus dan yang berisiko gagal (Andika, Nurhakim, & Andas, 2025).

Beberapa penelitian sebelumnya telah menerapkan metode machine learning, seperti Logistic Regression, Decision Tree, dan Random Forest, untuk memprediksi performa akademik siswa (Domladovac, 2021; Matzavela & Alepis, 2021). Penelitian-penelitian tersebut umumnya berfokus pada peningkatan akurasi model prediksi, dengan tujuan mengidentifikasi siswa yang berisiko gagal secara lebih tepat. Meskipun demikian, aspek interpretabilitas model, yaitu kemampuan untuk memahami kontribusi masing-masing fitur terhadap hasil prediksi, sering kali belum menjadi fokus utama dalam kajian-kajian tersebut.

Padahal, interpretabilitas memiliki peran penting dalam konteks pendidikan, karena memungkinkan pengambil keputusan untuk memahami faktor-faktor kunci yang memengaruhi performa akademik siswa dan merancang intervensi yang tepat sasaran. Tanpa interpretasi yang jelas, hasil prediksi model cenderung sulit dimanfaatkan secara praktis, meskipun memiliki tingkat akurasi yang tinggi. Beberapa studi menekankan bahwa informasi seperti kehadiran, jumlah kegagalan sebelumnya, dan dukungan keluarga merupakan indikator penting yang seharusnya dapat dijelaskan secara eksplisit oleh model prediksi (Lu, Swisher, Chung, Jaffray, & Sidey-Gibbons, 2023).

Namun, masih terdapat celah penelitian yang signifikan, yaitu terbatasnya studi yang mengintegrasikan performa prediksi yang memadai dengan analisis interpretabilitas fitur pada dataset pendidikan yang relatif kecil dan kompleks. Sebagian besar penelitian sebelumnya belum secara khusus menyoroti bagaimana informasi feature importance dapat dimanfaatkan sebagai dasar rekomendasi intervensi pendidikan yang aplikatif. Oleh karena itu, penelitian ini mengusulkan penggunaan algoritma XGBoost pada dataset Student Performance dari UCI, yang mencakup variabel akademik, sosial, dan demografis siswa sekolah menengah. XGBoost dipilih karena kemampuannya menangani dataset berukuran kecil dengan struktur kompleks, sekaligus menyediakan informasi feature importance yang dapat diinterpretasikan secara langsung untuk mendukung pengambilan keputusan pendidikan (Zhang, Jia, & Shang, 2022).

Tujuan penelitian ini adalah membangun model prediksi kelulusan siswa menggunakan XGBoost dengan performa yang memadai, mengidentifikasi fitur-fitur paling berpengaruh melalui analisis feature importance, dan memberikan rekomendasi intervensi pendidikan berdasarkan interpretasi fitur tersebut. Penelitian ini mencoba menjawab beberapa pertanyaan utama: seberapa akurat model XGBoost dalam memprediksi kelulusan siswa pada dataset Student Performance, fitur-fitur mana yang paling berpengaruh terhadap prediksi kelulusan, dan bagaimana informasi dari fitur-fitur penting dapat digunakan untuk mendukung pengambilan keputusan pendidikan atau merancang intervensi bagi siswa berisiko gagal.

Dengan fokus pada interpretabilitas dan pemahaman faktor-faktor penentu performa akademik, penelitian ini diharapkan memberikan kontribusi praktis bagi institusi pendidikan dan membangun dasar bagi penelitian lanjutan di bidang prediksi performa akademik siswa.

METODE PENELITIAN

Dataset

Penelitian ini menggunakan Student Performance Data Set dari UCI Machine Learning Repository, yang mencakup 395 siswa dari dua sekolah menengah di Portugal. Dataset berisi 33 variabel akademik, sosial, dan demografis, dengan target kelulusan pass/fail yang diturunkan dari nilai akhir G3. Siswa diberi label *pass* jika $G3 \geq 10$ dan *fail* jika $G3 < 10$. Fitur input mencakup berbagai aspek, seperti jumlah kegagalan sebelumnya, dukungan pendidikan tambahan, waktu belajar, absensi, partisipasi kegiatan ekstrakurikuler, kondisi kesehatan, hubungan keluarga, status wali, serta data demografis seperti usia, jenis kelamin, dan alamat. Nilai G1, G2, dan G3 asli dihapus sebelum pemodelan untuk mencegah kebocoran data (*data leakage*), sedangkan G3 digunakan untuk membuat target pass/fail.

Tabel 1 Deskripsi Fitur Dataset Student Performance

No	Fitur	Keterangan
1	school	Sekolah siswa (GP atau MS)
2	sex	Jenis kelamin siswa
3	age	Usia siswa
4	address	Tipe alamat (kota/desa)
5	famsize	Ukuran keluarga
6	Pstatus	Status orang tua (tinggal bersama/terpisah)
7	Medu	Pendidikan ibu
8	Fedu	Pendidikan ayah
9	Mjob	Pekerjaan ibu
10	Fjob	Pekerjaan ayah
11	reason	Alasan memilih sekolah
12	guardian	Wali yang bertanggung jawab
13	traveltime	Waktu tempuh dari rumah ke sekolah
14	studytime	Waktu belajar per minggu
15	failures	Jumlah kegagalan sebelumnya
16	schoolsup	Dukungan pendidikan tambahan
17	famsup	Dukungan keluarga
18	paid	Kursus tambahan dibayar
19	activities	Partisipasi dalam kegiatan ekstrakurikuler
20	nursery	Pernah bersekolah taman kanak-kanak
21	higher	Keinginan melanjutkan ke pendidikan tinggi
22	internet	Akses internet di rumah
23	romantic	Sedang dalam hubungan asmara
24	famrel	Kualitas hubungan keluarga
25	freetime	Waktu luang di luar sekolah
26	goout	Frekuensi keluar bersama teman
27	Dalc	Konsumsi alkohol pada hari kerja
28	Walc	Konsumsi alkohol pada akhir pekan
29	health	Kondisi kesehatan
30	absences	Jumlah ketidakhadiran
31	G1	Nilai pertama (semester awal) – dihapus sebelum modeling
32	G2	Nilai kedua (semester menengah) – dihapus sebelum modeling
33	G3	Nilai akhir – dijadikan target pass/fail dan dihapus dari fitur input

Preprocessing Data

Semua fitur kategorikal diencode menggunakan Label Encoder agar dapat digunakan dalam model XGBoost. Dataset memiliki distribusi kelas yang tidak seimbang, dengan kelas fail lebih sedikit dibanding pass. Oleh karena itu, evaluasi model menekankan metrik seperti recall dan F1-score untuk kelas minoritas. Dataset dibagi menjadi subset train dan test dengan perbandingan 80:20 secara stratified, sehingga proporsi kelas pass dan fail tetap terjaga di kedua subset.

Model Prediksi

Model yang digunakan adalah XGBoost Classifier, karena kemampuannya menangani

dataset kecil, fitur campuran (numerik dan kategorikal), serta menyediakan informasi feature importance yang interpretabel (Sukmawati, Masruriyah, Juwita, Tejayanda, & Nurmawanti, 2024). Model dilatih menggunakan subset train dan diuji pada subset test untuk mengevaluasi performa prediksi. Hyperparameter yang digunakan adalah $n_estimators=200$, $max_depth=10$, $learning_rate=0.01$, $eval_metric='logloss'$, dan $random_state=80$.

Evaluasi Model

Evaluasi dilakukan menggunakan metrik accuracy untuk mengukur persentase prediksi benar secara keseluruhan, serta precision, recall, dan F1-score untuk menilai performa prediksi per kelas, dengan fokus pada kelas minoritas fail. Macro average dihitung untuk memberikan gambaran keseimbangan prediksi antar kelas, sedangkan weighted average memperhitungkan proporsi masing-masing kelas (Fauzan, Vitianingsih, Cahyono, Maukar, & Suprio, 2025).

Analisis Feature Importance

Analisis feature importance dilakukan untuk menilai kontribusi relatif tiap fitur terhadap prediksi kelulusan, sehingga dapat digunakan sebagai dasar interpretasi dan rekomendasi intervensi pendidikan (Rajbahadur, Wang, Oliva, Kamei, & Hassan, 2022). Di bagian Metode, tabel ini hanya menyebutkan nama dan jenis fitur, sedangkan ranking dan nilai feature importance akan ditampilkan lengkap di Hasil & Pembahasan.

Keterbatasan Metode

Dataset relatif kecil dan kelas minoritas fail lebih sedikit dibanding pass, sehingga model cenderung memiliki recall rendah untuk siswa yang berisiko gagal. Analisis feature importance memberikan interpretasi global, tetapi tidak menangkap interaksi lokal antar fitur untuk setiap siswa. Studi ini hanya menggunakan satu dataset tunggal, sehingga generalisasi hasil ke populasi berbeda perlu diuji lebih lanjut.

HASIL DAN PEMBAHASAN

Setelah melatih model XGBoost pada dataset Student Performance, evaluasi dilakukan menggunakan subset test (20% dari dataset). Hasil evaluasi diperoleh sebagai berikut:

Tabel 2 Hasil Evaluasi Model XGBoost

Metric	fail	pass	Macro Avg	Weighted Avg
Precision	0.59	0.74	0.67	0.69
Recall	0.38	0.87	0.63	0.71
F1-score	0.47	0.80	0.63	0.69
Support	26	53	79	79

Evaluasi menunjukkan bahwa accuracy keseluruhan model mencapai 0.71, sedangkan Macro F1-score sebesar 0.63. Model cukup baik dalam mengklasifikasikan kelas mayoritas pass, terlihat dari recall 0.87 dan F1-score 0.80. Namun, kinerja pada kelas minoritas fail lebih rendah, dengan recall hanya 0.38, sehingga sebagian siswa yang berisiko gagal tidak terdeteksi. Hal ini sejalan dengan karakter dataset yang imbalanced, di mana siswa berstatus fail lebih sedikit. Meskipun demikian, Macro F1-score 0.63 menunjukkan model masih cukup seimbang antara kelas mayoritas dan minoritas, walaupun sensitivitas terhadap kelas fail perlu diperhatikan lebih lanjut.

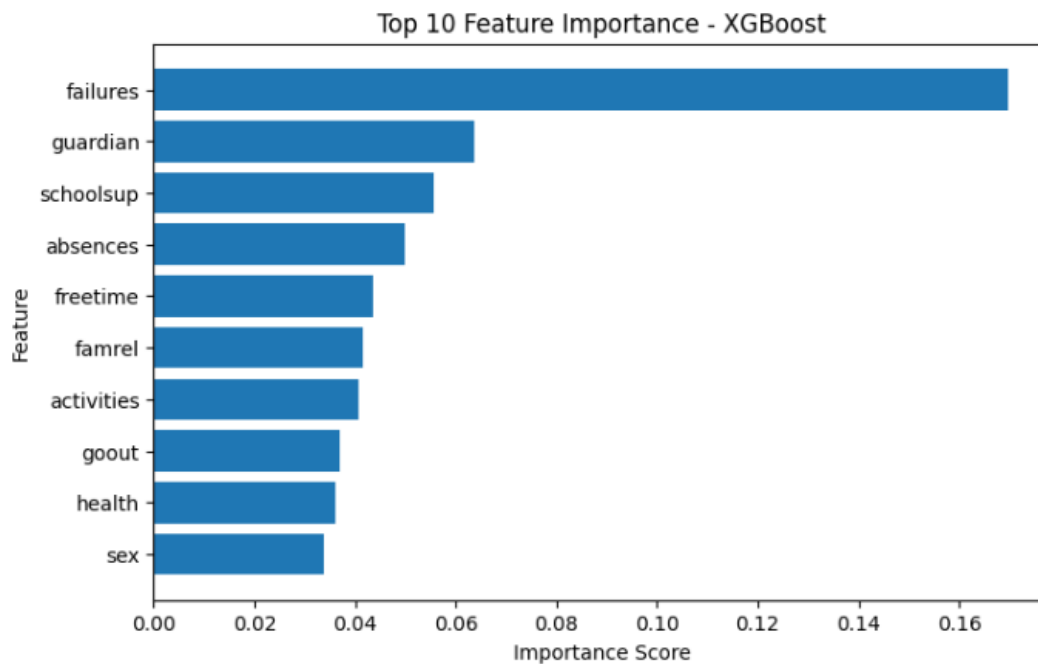
Kesimpulan sementara: XGBoost memberikan prediksi yang memadai untuk siswa mayoritas, sambil mempertahankan kemampuan interpretabilitas, sehingga memungkinkan analisis lebih lanjut terhadap faktor-faktor yang memengaruhi kelulusan. Namun, perhatian lebih diperlukan untuk menangani siswa berisiko gagal.

Top 10 fitur yang paling berpengaruh terhadap prediksi XGBoost ditampilkan pada Tabel 3 dan divisualisasikan pada Gambar 1.

Tabel 3 Top 10 Feature Importance (angka)

No	Feature	Importance
1	failures	0.169710
2	guardian	0.063610
3	schoolsup	0.055597

No	Feature	Importance
4	absences	0.049924
5	freetime	0.043577
6	famrel	0.041650
7	activities	0.040602
8	goout	0.037057
9	health	0.036188
10	sex	0.033752



Gambar 1: Visualisasi bar chart Top 10 Feature Importance

Analisis feature importance menunjukkan bahwa failures, atau jumlah kegagalan sebelumnya, merupakan faktor paling kritis, mengindikasikan siswa yang pernah gagal memiliki risiko lebih tinggi untuk gagal kembali. Temuan ini konsisten dengan literatur sebelumnya (Cortez & Silva, 2008). Selain itu, guardian (status wali atau orang tua) dan schoolsup (dukungan pendidikan tambahan) muncul sebagai faktor sosial signifikan, menekankan pentingnya dukungan keluarga dan bimbingan tambahan bagi siswa sekolah menengah. Fitur lain, seperti absences (jumlah absen) dan freetime (waktu luang), juga berperan, menunjukkan bahwa disiplin dan manajemen waktu menjadi indikator penting performa siswa. Fitur-fitur tambahan seperti famrel, activities, goout, health, dan sex memberikan kontribusi lebih kecil, tetapi tetap relevan untuk merancang intervensi pendidikan yang holistik.

Visualisasi Top 10 fitur memperlihatkan secara jelas pengaruh relatif tiap fitur terhadap prediksi, sehingga institusi pendidikan dapat memprioritaskan intervensi berdasarkan faktor teratas, misalnya membantu siswa dengan kegagalan sebelumnya, meningkatkan dukungan wali, atau meninjau program sekolah tambahan.

Berdasarkan hasil evaluasi dan analisis feature importance, model XGBoost menunjukkan accuracy yang memadai meskipun recall pada kelas minoritas fail relatif rendah. Analisis feature importance memungkinkan interpretasi faktor-faktor utama yang memengaruhi kelulusan siswa, sehingga institusi pendidikan dapat menargetkan intervensi secara lebih efektif berdasarkan fitur paling signifikan. Kombinasi antara metrik evaluasi dan feature importance memberikan pemahaman praktis, bukan sekadar akurasi model, sehingga hasil penelitian ini lebih aplikatif untuk pengambilan keputusan pendidikan.

Beberapa studi sebelumnya menggunakan algoritma Decision Tree, Random Forest, dan SVM untuk prediksi performa akademik atau dropout siswa, dengan performa bervariasi tergantung karakteristik dataset (Akhter et al., 2021; Wang et al., 2022). Secara umum, XGBoost menunjukkan performa kompetitif atau lebih tinggi dibanding Random Forest pada dataset dengan fitur campuran dan kelas minoritas, terutama dalam hal akurasi keseluruhan dan

kemampuan menangani data imbalanced. Perbedaan utama penelitian ini adalah fokus pada interpretabilitas fitur, sehingga hasil prediksi dapat digunakan langsung untuk mendukung pengambilan keputusan pendidikan.

Keterbatasan penelitian ini mencakup ukuran dataset yang relatif kecil serta performa model yang masih rendah pada kelas minoritas siswa berisiko gagal, yang tercermin dari nilai recall yang rendah. Kondisi ini berimplikasi pada meningkatnya potensi false negative, yaitu siswa yang sebenarnya berisiko gagal namun diprediksi sebagai lulus oleh model. Dalam konteks pendidikan, false negative memiliki implikasi yang signifikan karena siswa tersebut berpotensi tidak mendapatkan intervensi dini, seperti bimbingan tambahan atau pendampingan akademik, sehingga risiko kegagalan akademik tetap tinggi. Selain itu, analisis feature importance yang digunakan dalam penelitian ini bersifat global dan belum mampu menangkap variasi serta interaksi lokal antar karakteristik siswa secara individual. Oleh karena itu, penelitian lanjutan disarankan untuk mengeksplorasi teknik penanganan data tidak seimbang, seperti SMOTE atau penyesuaian threshold klasifikasi, serta menggunakan dataset yang lebih besar dan beragam guna meningkatkan sensitivitas model terhadap siswa berisiko gagal dan meminimalkan dampak false negative dalam penerapan nyata.

KESIMPULAN

Berdasarkan penelitian pada siswa sekolah menengah menggunakan dataset Student Performance, dapat disimpulkan bahwa model XGBoost efektif untuk memprediksi kelulusan siswa, dengan accuracy 0.71 dan Macro F1-score 0.63. Model ini mampu menangkap pola penting dari data akademik dan sosial, meskipun recall pada kelas minoritas fail lebih rendah (0.38). Analisis feature importance memberikan interpretasi yang jelas, dengan *failures* (jumlah kegagalan sebelumnya) sebagai faktor paling kritis, diikuti oleh *guardian* (status wali/orang tua) dan *schoolsup* (dukungan pendidikan tambahan). Fitur lain, seperti *absences*, *freetime*, dan *activities*, juga memberikan kontribusi signifikan. Kombinasi antara performa prediksi dan interpretabilitas fitur menjadikan model ini tidak hanya berguna untuk evaluasi akademik, tetapi juga sebagai alat bantu pengambilan keputusan intervensi pendidikan. Untuk penelitian lanjutan, disarankan meningkatkan deteksi kelas minoritas melalui teknik balancing atau penyesuaian threshold, serta menggunakan dataset yang lebih besar untuk meningkatkan generalisasi model. Informasi dari fitur-fitur penting seperti *failures*, *guardian*, dan *schoolsup* dapat dijadikan dasar pengembangan program bimbingan atau dukungan tambahan bagi siswa yang berisiko gagal, sehingga model tidak hanya memberikan prediksi, tetapi juga kontribusi praktis bagi intervensi pendidikan. Dengan penerapan saran-saran tersebut, model prediksi performa akademik tidak hanya akurat, tetapi juga dapat memberikan dampak nyata bagi perencanaan pendidikan dan pengambilan keputusan institusi.

REFERENSI

- Alfarizi, M. B., Witanti, W., & Komarudin, A. (2025). Prediksi Kinerja Akademik Siswa Bimbingan Belajar Menggunakan Algoritma Extreme Gradient Boosting (XGBoost). *Building of Informatics, Technology and Science (BITS)*, 7(1), 584–594.
- Andika, A. W., Nurhakim, L., & Andas, N. H. (2025). Penggunaan Deep Learning Untuk Memprediksi Kinerja Akademik Dan Memberi Dukungan Yang Tepat Bagi Siswa. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, 4(7), 1647–1664.
- Domladovac, M. (2021). Comparison of Neural Network with Gradient Boosted Trees, Random Forest, Logistic Regression and SVM in predicting student achievement. *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, 211–216. IEEE. <https://doi.org/10.23919/MIPRO52101.2021.9596684>
- Fauzan, R., Vitianingsih, A. V., Cahyono, D., Maukar, A. L., & Suprio, Y. A. B. (2025). Penerapan Algoritma Klasifikasi pada Machine Learning untuk Deteksi Phishing. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 531–540. <https://doi.org/10.57152/malcom.v5i2.1968>
- Lu, S.-C., Swisher, C. L., Chung, C., Jaffray, D., & Sidey-Gibbons, C. (2023). On the importance of interpretable machine learning predictions to inform clinical decision making in oncology. *Frontiers in Oncology*, 13. <https://doi.org/10.3389/fonc.2023.1129380>
- Matzavela, V., & Alepis, E. (2021). Decision tree learning through a Predictive Model for Student Academic Performance in Intelligent M-Learning environments. *Computers and Education: Artificial Intelligence*, 2, 100035. <https://doi.org/10.1016/j.caeai.2021.100035>

- Rajbahadur, G. K., Wang, S., Oliva, G. A., Kamei, Y., & Hassan, A. E. (2022). The Impact of Feature Importance Methods on the Interpretation of Defect Classifiers. *IEEE Transactions on Software Engineering*, 48(7), 2245–2261. <https://doi.org/10.1109/TSE.2021.3056941>
- Ravindra, P., Yogita, K., Shrikrishna, Y., Lalita, P., & Aniket, S. (2024). Academic Performance Indicators For Students Of Engineering College. *International Journal of Creative Research Thoughts (IJCRT)*, 12(4), 134–139.
- Simbolon, I., Aditya, P., & Br Purba, E. (2025). Prediksi Performa Akademik Siswa Berdasarkan Kehadiran dan Aktivitas E-Learning Menggunakan Algoritma Decision Tree. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 4899–4910. <https://doi.org/10.31004/riggs.v4i2.1352>
- Sukmawati, C. E., Masruriyah, A. F. N., Juwita, A. R., Tejayanda, R. D., & Nurmayanti, T. (2024). Evaluasi Kinerja Algoritma AdaBoost dan XGBoost Menggunakan Dataset Penyakit Obesitas Pada Populasi Dewasa. *Jambura Journal of Informatics*, 6(2), 101–111.
- Suleiman, I. B., Okunade, O. A., Dada, E. G., & Ezeanya, U. C. (2024). Key factors influencing students' academic performance. *Journal of Electrical Systems and Information Technology*, 11(1), 41. <https://doi.org/10.1186/s43067-024-00166-w>
- Wijaya, V. Y. D., & Brotosaputro, G. (2025). Penerapan Data Mining Dalam Prediksi Kinerja Akademik Mahasiswa Menggunakan Algoritma Machine Learning. *Jurnal Infomedia: Teknik Informatika, Multimedia, Dan Jaringan*, 10(2), 134–142.
- Yanuarini Nur Sukmaningtyas, Makhfuddin Akbar, R., & Rohma Utami Asyafiiyah, G. (2024). Penerapan Predictive Analytics untuk Analisis Faktor-faktor yang Mempengaruhi Performa Akademik Siswa. *Arcitech: Journal of Computer Science and Artificial Intelligence*, 4(2), 127–145. <https://doi.org/10.29240/arcitech.v4i2.12048>
- Zhang, P., Jia, Y., & Shang, Y. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, 18(6), 155013292211069. <https://doi.org/10.1177/15501329221106935>