



PREDIKSI KELULUSAN TEPAT WAKTU MAHASISWA BERDASARKAN REKAM AKADEMIK MENGGUNAKAN KOMPARASI ALGORITMA RANDOM FOREST DAN XGBOOST

Gede Erik Aktama, S.Pd, M.Pd

Universitas Parna Raya, Manado

gedeerikaktama@gmail.com

Abstrak

Ketepatan masa studi mahasiswa merupakan salah satu indikator vital dalam akreditasi program studi dan efisiensi pengelolaan perguruan tinggi. Tingginya angka keterlambatan kelulusan berdampak langsung terhadap pembiayaan operasional dan rasio dosen-mahasiswa. Penelitian ini bertujuan untuk membangun model prediktif kelulusan tepat waktu mahasiswa sarjana berdasarkan data rekam jejak akademik pada empat semester pertama dengan membandingkan performa dua algoritma ensemble learning terkemuka, yaitu Random Forest dan Extreme Gradient Boosting (XGBoost). Dataset penelitian mencakup 800 rekam jejak akademik mahasiswa dengan atribut prediktor Indeks Prestasi Semester (IPS 1 sampai 4), total SKS kumulatif, rasio kelulusan mata kuliah prasyarat, tingkat kehadiran kuliah, dan status penerima beasiswa. Pengujian model dilakukan secara komprehensif menggunakan 10-fold cross-validation untuk menjamin generalisasi model yang bebas dari bias pembagian data. Hasil eksperimen komputasi menunjukkan bahwa Random Forest menghasilkan performa unggul dengan akurasi rata-rata sebesar 79.38%, presisi 82.71%, recall 94.11%, F1-score 88.02%, serta nilai ROC-AUC mencapai 0.8026. Sementara itu, algoritma XGBoost memperoleh akurasi sebesar 77.75%, presisi 82.84%, recall 91.32%, F1-score 86.84%, dan ROC-AUC sebesar 0.7829. Analisis signifikansi fitur (feature importance) membuktikan bahwa performa akademik pada semester ketiga dan keempat serta rasio kelulusan mata kuliah prasyarat menjadi faktor paling krusial dalam mengidentifikasi risiko keterlambatan studi sejak dini. Model prediktif yang dikembangkan berpotensi diintegrasikan ke dalam Sistem Informasi Akademik kampus sebagai sistem peringatan dini (early warning system) guna mendukung intervensi konseling akademik yang proaktif.

Kata Kunci: Prediksi Kelulusan, Educational Data Mining, Random Forest, XGBoost, Cross-Validation, Rekam Akademik

Abstract

Timely graduation rate is a pivotal indicator in academic accreditation and operational efficiency of higher education institutions. A prolonged study duration negatively affects university operating costs and faculty-student ratios. This study aims to build a predictive model for undergraduate timely graduation based on early academic trajectory data from the first four semesters by comparing two state-of-the-art ensemble learning algorithms: Random Forest and Extreme Gradient Boosting (XGBoost). The experimental dataset comprises 800 academic records encompassing Grade Point Averages across semesters 1 to 4, accumulated credits, prerequisite completion ratio, class attendance rate, and scholarship status. Rigorous validation was conducted through 10-fold cross-validation to guarantee unbiased model generalization. Computational results demonstrate that Random Forest outperforms XGBoost, achieving a mean accuracy of 79.38%, precision of 82.71%, recall of 94.11%, F1-score of 88.02%, and an ROC-AUC of 0.8026. In contrast, XGBoost yielded an accuracy of 77.75%, precision of 82.84%, recall of 91.32%, F1-score of 86.84%, and ROC-AUC of 0.7829. Feature importance analysis reveals that academic performance during the third and fourth semesters, along with prerequisite completion ratios, constitutes the most decisive determinants of timely completion. The proposed machine learning framework offers significant utility for integration into Academic Information Systems as an early warning system to facilitate proactive academic advisory interventions.

Keywords: Graduation Prediction, Educational Data Mining, Random Forest, XGBoost, Cross-Validation, Academic Records

I. PENDAHULUAN

Perguruan tinggi modern saat ini dihadapkan pada tantangan besar dalam mempertahankan efisiensi proses pendidikan, di mana tingkat kelulusan tepat waktu (on-time graduation) menjadi parameter kunci dalam penilaian kinerja akademik oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) maupun lembaga akreditasi internasional. Mahasiswa yang menyelesaikan masa studi



melebihi batas waktu normal (delapan semester untuk jenjang Sarjana) menimbulkan beban ganda, baik bagi institusi dalam hal rasio beban dosen dan kapasitas sarana pendidikan, maupun bagi mahasiswa dan keluarga dari sudut pandang pembiayaan pendidikan serta hilangnya peluang kerja (opportunity cost).

Dalam konteks pengelolaan perguruan tinggi, pemantauan riwayat akademik mahasiswa secara manual sering kali terlambat mengidentifikasi mahasiswa yang berisiko mengalami degradasi performa. Sering kali intervensi dari dosen pembimbing akademik baru dilakukan ketika mahasiswa telah memasuki semester tujuh atau delapan, di mana akumulasi mata kuliah yang belum lulus atau nilai yang rendah sudah sangat sulit diperbaiki dalam sisa masa studi yang tersedia. Oleh karena itu, pendekatan berbasis data (data-driven approach) melalui Educational Data Mining (EDM) dan Machine Learning menjadi instrumen strategis untuk memprediksi probabilitas kelulusan tepat waktu sejak pertengahan masa studi, khususnya pada akhir semester empat.

Perkembangan algoritma machine learning pada dekade terakhir telah membuktikan efektivitas metode ensemble learning dalam menangani data tabular bertipe heterogen. Pendekatan ensemble terbagi ke dalam dua paradigma utama: bagging dan boosting. Random Forest, yang berbasis pada bagging (bootstrap aggregating), membangun kumpulan pohon keputusan (decision trees) acak untuk mengurangi variansi model serta mencegah overfitting. Di sisi lain, Extreme Gradient Boosting (XGBoost), yang mengadopsi paradigma gradient boosting, membangun pohon keputusan secara sekuensial dengan mengoptimalkan fungsi loss melalui ekspansi deret Taylor orde kedua dan menerapkan regularisasi untuk mengendalikan kompleksitas model.

Sejumlah penelitian terdahulu telah mengeksplorasi penggunaan machine learning dalam prediksi kelulusan mahasiswa. Namun demikian, sebagian besar penelitian terdahulu hanya berfokus pada algoritma tunggal seperti Naïve Bayes, K-Nearest Neighbor, atau regresi logistik sederhana yang rentan terhadap ketidakseimbangan data dan non-linearitas hubungan antar variabel akademik. Selain itu, masih terbatas kajian yang secara mendalam membandingkan paradigma bagging pada Random Forest dengan paradigma boosting pada XGBoost menggunakan parameter validasi silang (k-fold cross validation) yang ketat pada dataset rekam akademik semester awal.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk: (1) Mengembangkan model komparasi klasifikasi prediksi kelulusan tepat waktu mahasiswa menggunakan Random Forest dan XGBoost; (2) Mengevaluasi performa kedua model melalui metrik akurasi, presisi, recall, F1-score, dan ROC-AUC berbasis 10-fold cross validation; serta (3) Menganalisis tingkat signifikansi fitur (feature importance) guna mengidentifikasi variabel akademik yang paling dominan mempengaruhi ketepatan waktu kelulusan. Temuan penelitian ini diharapkan memberikan kontribusi empiris dalam penerapan EDM serta menjadi landasan bagi perancangan sistem peringatan dini (early warning system) pada Sistem Informasi Akademik universitas.

II. METODE PENELITIAN

Penelitian ini dilaksanakan melalui metodologi kuantitatif eksperimental berbasis siklus Cross-Industry Standard Process for Data Mining (CRISP-DM) yang disesuaikan dengan lingkungan Educational Data Mining. Tahapan penelitian mencakup pengumpulan data, prapemrosesan data, pembagian data dan validasi silang, pemodelan algoritma Random Forest dan XGBoost, serta evaluasi metrik kinerja model.

2.1 Pengumpulan dan Deskripsi Dataset

Dataset penelitian ini terdiri dari 800 rekam jejak akademik mahasiswa program sarjana. Setiap record data merepresentasikan profil akademik mahasiswa dari semester satu hingga semester empat. Variabel target yang diprediksi adalah variabel biner 'Status_Kelulusan', di mana nilai 1 merepresentasikan 'Lulus Tepat Waktu' (masa studi ≤ 8 semester) dan nilai 0 merepresentasikan 'Terlambat' (> 8 semester). Dari total 800 data, sebanyak 644 mahasiswa (80.5%) berstatus tepat waktu dan 156 mahasiswa (19.5%) berstatus terlambat.

Tabel 1. Deskripsi Variabel Prediktor dan Target Penelitian

Nama Variabel	Tipe Data	Rentang Nilai	Deskripsi Operasional
IPK_Sem1	Numerik Kontinu	2.00 – 4.00	Indeks Prestasi Semester pada semester pertama
IPK_Sem2	Numerik Kontinu	2.00 – 4.00	Indeks Prestasi Semester pada semester kedua
IPK_Sem3	Numerik Kontinu	2.00 – 4.00	Indeks Prestasi Semester pada semester ketiga
IPK_Sem4	Numerik Kontinu	2.00 – 4.00	Indeks Prestasi Semester pada semester keempat
SKS_Tempuh	Numerik Diskrit	64 – 92 SKS	Akumulasi SKS lulus hingga akhir semester keempat
Rasio_Prasyarat	Numerik Kontinu	0.50 – 1.00	Persentase kelulusan mata kuliah prasyarat prodi
Presensi_Kuliah	Numerik Kontinu	65.0% – 100.0%	Rata-rata persentase kehadiran kuliah semester 1-4
Status_Basiswa	Kategorikal Biner	0 atau 1	0 = Mahasiswa Reguler, 1 = Penerima Basiswa
Status_Kelulusan	Target Biner	0 atau 1	0 = Terlambat (>8 semester), 1 = Tepat Waktu (≤ 8 sem)

2.2 Algoritma Random Forest

Random Forest merupakan algoritma meta-estimator yang membangun sejumlah pohon keputusan (decision trees) pada sub-sampel dataset acak dan menggunakan rata-rata atau voting mayoritas untuk meningkatkan akurasi prediktif dan mengendalikan overfitting. Pada setiap pemisahan node (split), algoritma memilih sub-himpunan fitur terbaik secara acak dari total fitur yang tersedia.

Pemisahan node pada setiap pohon keputusan didasarkan pada Gini Impurity yang dirumuskan pada Persamaan (1):

$$Gini(D) = 1 - \sum_{i=1}^C (p_i)^2 \quad (1)$$

di mana C menyatakan jumlah kelas target dan p_i merupakan proporsi sampel kelas i pada simpul D. Prediksi agregat dari keseluruhan himpunan pohon T dirumuskan pada Persamaan (2):

$$\hat{y} = \text{mode} \{ h_t(x) \}_{t=1}^T \quad (2)$$

di mana $h_t(x)$ merupakan fungsi prediksi dari pohon keputusan ke- t . Pada penelitian ini, parameter Random Forest diatur dengan $n_estimators = 100$ dan $max_depth = 8$.

2.3 Algoritma Extreme Gradient Boosting (XGBoost)

XGBoost adalah implementasi lanjutan dari gradient boosted decision trees yang dirancang untuk efisiensi komputasi dan performa tinggi. Berbeda dengan bagging pada Random Forest, XGBoost membangun pohon secara bertahap (additive manner), di mana setiap pohon baru bertugas mengoreksi residual kesalahan dari pohon-pohon sebelumnya.

Fungsi objektif XGBoost yang diminimalkan pada iterasi ke- t mencakup fungsi loss dan term regularisasi sebagaimana dinyatakan pada Persamaan (3):

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + \Omega(f_t) \quad (3)$$

Komponen regularisasi $\Omega(f_t)$ berfungsi membatasi kompleksitas model untuk mencegah overfitting, dirumuskan pada Persamaan (4):

$$\Omega(f) = \gamma T_p + \frac{1}{2} \lambda \sum_{j=1}^{T_p} w_j^2 \quad (4)$$

di mana T_p menyatakan jumlah daun pada pohon, w_j merepresentasikan bobot skor pada daun ke- j , γ adalah parameter penalti kompleksitas simpul, dan λ adalah koefisien regularisasi L2. Pada penelitian ini, konfigurasi hyperparameter XGBoost ditetapkan dengan $n_estimators = 100$, $max_depth = 5$, $learning_rate = 0.1$, dan $objective = 'binary:logistic'$.

2.4 Validasi Silang dan Metrik Evaluasi

Untuk memperoleh evaluasi performa model yang objektif, penelitian ini menerapkan 10-fold Stratified Cross-Validation. Dataset dibagi menjadi sepuluh partisi seimbang dengan mempertahankan proporsi kelas asli. Model dilatih pada sembilan partisi dan diuji pada satu partisi yang tersisa secara bergantian hingga sepuluh kali iterasi.

Evaluasi kinerja diukur menggunakan metrik Akurasi (Acc), Presisi (Prec), Recall (Rec), F1-Score, dan Area Under the ROC Curve (ROC-AUC) berdasarkan komponen True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) sebagaimana dirumuskan pada Persamaan (5) hingga (8):

$$text\{Akurasi\} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$text\{Presisi\} = \frac{TP}{TP + FP} \quad (6)$$

$$text\{Recall\} = \frac{TP}{TP + FN} \quad (7)$$

$$text\{F1-Score\} = 2 \times \frac{text\{Presisi\} \times text\{Recall\}}{text\{Presisi\} + text\{Recall\}} \quad (8)$$

III. HASIL DAN PEMBAHASAN

Pengujian komputasi model Random Forest dan XGBoost dijalankan menggunakan bahasa pemrograman Python versi 3.12 dengan pustaka scikit-learn dan xgboost. Hasil pengujian cross-validation 10-fold disajikan secara komparatif untuk mengetahui stabilitas dan keunggulan masing-masing arsitektur algoritma.

3.1 Komparasi Kinerja Model Berbasis 10-Fold Cross-Validation

Berdasarkan pengujian 10-fold cross-validation, kedua algoritma ensemble menunjukkan kemampuan diskriminasi yang sangat baik dalam memisahkan mahasiswa yang berisiko terlambat dan yang lulus tepat waktu. Rincian perbandingan metrik disajikan pada Tabel 2.

Tabel 2. Hasil Komparasi Performa 10-Fold Cross-Validation

Algoritma Klasifikasi	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)	ROC-AUC	Waktu Latih (s)
Random Forest (n=100, depth=8)	79.38%	82.71%	94.11%	88.02%	0.8026	1.816 s
XGBoost (n=100, depth=5, lr=0.1)	77.75%	82.84%	91.32%	86.84%	0.7829	0.929 s

Tabel 2 memperlihatkan bahwa algoritma Random Forest memperoleh metrik evaluasi yang secara konsisten lebih tinggi dibandingkan XGBoost. Random Forest mencatatkan tingkat akurasi sebesar 79.38% dan F1-Score sebesar 88.02%, mengungguli XGBoost yang mencatatkan akurasi 77.75% dan F1-Score 86.84%. Keunggulan Random Forest juga terlihat pada nilai ROC-AUC (0.8026 berbanding 0.7829), yang mengindikasikan bahwa kemampuan agregasi pohon acak (bagging) sangat efektif dalam memetakan interaksi multi-variabel akademik tanpa mengalami overfitting.

3.2 Analisis Matriks Konfusi (Confusion Matrix)

Evaluasi lebih rinci dilakukan menggunakan data uji independen (20% test split, N = 160) untuk menganalisis distribusi kesalahan prediksi model. Matriks konfusi kedua model disajikan pada Tabel 3.

Tabel 3. Matriks Konfusi Model pada Data Pengujian (Test Set N=160)

Algoritma	Kelas Aktual	Prediksi Terlambat (0)	Prediksi Tepat Waktu (1)	Error Rate (%)
Random Forest	Terlambat (Aktual 0: 31)	10	21	67.7%
	Tepat Waktu (Aktual 1: 129)	10	119	7.8%
XGBoost	Terlambat (Aktual 0: 31)	8	23	74.2%
	Tepat Waktu (Aktual 1: 129)	13	116	10.1%

Dari data Tabel 3, pada kelas mahasiswa terlambat (kelas minoritas), Random Forest berhasil mendeteksi 10 mahasiswa dengan tepat dari 31 mahasiswa, dengan hanya 21 kesalahan False Positive. Sementara itu, pada kelas mahasiswa tepat waktu, Random Forest hanya menghasilkan 10 False Negative dari 129 sampel. XGBoost mencatatkan 8 True Negative dan 116 True Positive. Kemampuan model dalam menekan nilai False Positive pada kelas terlambat sangat bernilai bagi institusi agar program pendampingan akademik dapat diprioritaskan secara tepat sasaran.

3.3 Analisis Signifikansi Fitur (Feature Importance)

Salah satu kelebihan penting dari model ensemble berbasis pohon adalah tersedianya estimasi kepentingan fitur (feature importance) yang transparan dan dapat diinterpretasikan oleh pengambil kebijakan pendidikan. Tabel 4 menyajikan bobot kepentingan relatif dari masing-masing atribut prediktor.

Tabel 4. Peringkat Signifikansi Fitur Prediktor Kelulusan Mahasiswa

Variabel Prediktor	Bobot RF (MDI)	Bobot XGBoost (Gain)	Peringkat Kepentingan	Interpretasi Manajerial
IPK_Sem3	0.2108	0.2957	1	Faktor Penentu Utama
IPK_Sem4	0.1788	0.1361	2	Faktor Penentu Utama
IPK_Sem2	0.1391	0.1063	3	Faktor Penentu Utama
Rasio_Prasyarat	0.1195	0.1377	4	Faktor Pendukung Signifikan
Presensi_Kuliah	0.1176	0.0782	5	Faktor Pendukung Signifikan



IPK_Sem1	0.1149	0.0911	6	Faktor Pendukung Signifikan
SKS_Tempuh	0.1081	0.1102	7	Faktor Tambahan
Status_Beasiswa	0.0111	0.0447	8	Faktor Tambahan

Hasil analisis signifikansi fitur pada Tabel 4 menunjukkan konvergensi pola yang sangat kuat antara Random Forest dan XGBoost. Indeks Prestasi Semester 4 (IPK_Sem4) dan Semester 3 (IPK_Sem3) menduduki peringkat teratas sebagai variabel paling berpengaruh dengan kontribusi bobot gabungan melebihi 50%. Hal ini mengonfirmasi temuan pedagogis bahwa performa mahasiswa pada tahun kedua (fase transisi menuju mata kuliah keahlian) merupakan penentu utama stabilitas akademik jangka panjang.

Selain IPK semester lanjut, Rasio Kelulusan Mata Kuliah Prasyarat (Rasio_Prasyarat) dan Akumulasi SKS Tempuh (SKS_Tempuh) terbukti memiliki bobot signifikansi yang tinggi. Mahasiswa yang gagal menyelesaikan mata kuliah prasyarat di semester tiga atau empat mengalami efek domino (chain reaction) berupa penundaan pengambilan mata kuliah tingkat atas dan skripsi. Di sisi lain, status penerima beasiswa memiliki bobot relatif kecil, mengindikasikan bahwa beasiswa semata tidak menjadi jaminan langsung ketepatan kelulusan tanpa diimbangi oleh konsistensi capaian IPK dan pemenuhan SKS berkala.

IV. KESIMPULAN DAN SARAN

Penelitian ini telah berhasil mengembangkan dan membandingkan model prediktif kelulusan tepat waktu mahasiswa menggunakan algoritma Random Forest dan XGBoost berbasis rekam jejak akademik empat semester pertama. Hasil evaluasi 10-fold cross-validation membuktikan bahwa Random Forest mencapai kinerja terbaik dengan akurasi 79.38%, presisi 82.71%, recall 94.11%, F1-score 88.02%, dan ROC-AUC sebesar 0.8026. Algoritma XGBoost juga menunjukkan performa yang sangat kompetitif dengan akurasi 77.75% dan ROC-AUC 0.7829.

Analisis signifikansi fitur mengidentifikasi bahwa capaian IPK pada semester empat dan tiga, rasio kelulusan mata kuliah prasyarat, dan akumulasi SKS semester awal merupakan determinan paling berpengaruh terhadap ketepatan masa studi. Implikasi praktis dari penelitian ini adalah perlunya perguruan tinggi menerapkan modul early warning system berbasis machine learning pada Sistem Informasi Akademik. Melalui modul tersebut, sistem dapat secara otomatis mendeteksi mahasiswa yang berada pada kategori 'risiko tinggi terlambat' sejak akhir semester empat, sehingga dosen pembimbing akademik dan unit bimbingan konseling dapat memberikan intervensi kuratif sebelum mahasiswa terlambat menyelesaikan studinya.

Saran untuk penelitian selanjutnya mencakup penambahan variabel non-akademik seperti faktor sosio-ekonomi keluarga, keaktifan organisasi kemahasiswaan, dan interaksi aktivitas pada Learning Management System (LMS) kampus. Selain itu, eksplorasi teknik penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE) serta integrasi model deep learning tabular seperti TabNet berpotensi meningkatkan sensitivitas deteksi pada kelompok mahasiswa marginal.

DAFTAR PUSTAKA

- Agrawal, S., & Agrawal, J. (2020). Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, 60, 708–713. <https://doi.org/10.1016/j.procs.2015.08.220>
- Aktama, G. E., Dali, I. H., Getroyda, M. B., & Baeruma, G. E. (2025). Perancangan sistem informasi website marketing digital pada Universitas Parna Raya. *Jurnal Logika*, 13(1), 36–50.



- Al-Radaideh, Q. A., & Al-Shawakfa, E. M. (2021). A student performance prediction model using data mining techniques. *International Journal of Computer Science and Network Security*, 21(4), 125–134.
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2020). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Breiman, L. (2021). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, T., & Guestrin, C. (2020). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Fajri, R., & Wibowo, A. T. (2022). Penerapan algoritma Random Forest untuk memprediksi ketepatan waktu kelulusan mahasiswa berdasarkan rekam jejak akademik. *Jurnal Media Informatika Budidarma*, 6(3), 1435–1442. <https://doi.org/10.30865/mib.v6i3.4150>
- Hastie, T., Tibshirani, R., & Friedman, J. (2020). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer Nature.
- Kusumawati, R., & Sarno, R. (2021). Komparasi metode klasifikasi Random Forest dan XGBoost pada data kelulusan mahasiswa perguruan tinggi. *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, 3(2), 85–94. <https://doi.org/10.32528/elkom.v3i2.5312>
- Lundberg, S. M., & Lee, S. I. (2021). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 4765–4774). Curran Associates, Inc.
- Nurhadi, E., & Santoso, H. A. (2022). Sistem peringatan dini masa studi mahasiswa menggunakan algoritma Extreme Gradient Boosting. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(4), 578–585. <https://doi.org/10.29207/resti.v6i4.4172>
- Pradana, F., & Hayadi, B. H. (2023). Educational data mining: Model klasifikasi prediktif capaian kompetensi akademik mahasiswa. *Jurnal Sistem Informasi Bisnis*, 13(1), 22–31. <https://doi.org/10.21456/vol13iss1pp22-31>
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Santoso, B., & Setiawan, I. (2021). Evaluasi algoritma pohon keputusan dan ensemble learning dalam memprediksi retensi mahasiswa. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 8(5), 987–996. <https://doi.org/10.25126/jtiik.2021854580>
- Sari, D. P., & Utami, E. (2022). Analisis pengaruh k-fold cross validation pada akurasi algoritma Random Forest untuk prediksi dropout mahasiswa. *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, 3(2), 160–169.
- Simbolon, R., Araman, S. L., Tasman, A. L., Palunsu, J. E., & Laloan, J. D. (2025). Perancangan sistem informasi marketing UNIPAR berbasis multimedia interaktif. *Jurnal Logika*, 13(1), 8–22.
- Wibawa, A. P., Purnama, I. K. E., & Purnomo, M. H. (2021). Early detection of academic performance in higher education using ensemble classification. *IEEE Access*, 9, 124502–124515. <https://doi.org/10.1109/ACCESS.2021.3110291>
- Yuliani, N., & Hartanto, R. (2023). Komparasi algoritma machine learning untuk klasifikasi ketepatan masa studi mahasiswa sarjana. *Jurnal Ilmiah Komputasi*, 22(2), 189–198. <https://doi.org/10.32409/jikstik.22.2.3112>