



IMPLEMENTASI MACHINE LEARNING UNTUK PREDIKSI RISIKO AKADEMIK MAHASISWA SEBAGAI EARLY WARNING SYSTEM PADA PERGURUAN TINGGI

Yuan Sa'dati¹, Nurul Umami²,

¹Teknologi Informasi, Universitas Qamarul Huda Badaruddin Bagu, Indonesia

²Teknologi Informasi, Institut Teknologi dan Kesehatan Aspirasi, Indonesia

Email : yuansaadati1610@gmail.com¹, nurulumamy27@gmail.com²

ABSTRAK

Penelitian ini bertujuan mengimplementasikan *Machine Learning* untuk memprediksi risiko akademik mahasiswa sebagai Early Warning System pada perguruan tinggi. Data yang digunakan meliputi IPK, IP semester, jumlah SKS, mata kuliah tidak lulus, mata kuliah mengulang, dan tingkat kehadiran. Tahapan penelitian meliputi *preprocessing*, seleksi fitur, pembagian data, pemodelan menggunakan *machine learning*, serta evaluasi dengan *accuracy*, *precision*, *recall*, dan *F1-score*. Model yang dihasilkan digunakan untuk mengklasifikasikan mahasiswa berdasarkan tingkat risiko akademik sehingga dapat membantu dosen dan program studi melakukan deteksi serta intervensi lebih awal. Hasil penelitian menunjukkan bahwa pendekatan *machine learning* berpotensi meningkatkan efektivitas pemantauan dan pengambilan keputusan akademik berbasis data.

Kata Kunci: *Machine Learning*, Risiko Akademik, *Early Warning System*, Mahasiswa, Prediksi.

ABSTRACT

This study aims to implement machine learning to predict students' academic risk, serving as an early warning system for higher education institutions. The data utilized includes cumulative GPA, semester GPA, total credit units (SKS), failed courses, repeated courses, and attendance rates. The research process encompasses data preprocessing, feature selection, data splitting, machine learning modeling, and performance evaluation using accuracy, precision, recall, and F1-score. The resulting model classifies students based on their academic risk level, thereby enabling lecturers and academic programs to facilitate early detection and intervention. The findings indicate that the machine learning approach has the potential to enhance the effectiveness of monitoring and data-driven academic decision-making.

Keywords: *Machine Learning*, Academic Risk, *Early Warning System*, Students, Prediction.

*Corresponding author: yuansaadati1610@gmail.com

© 2023 TAMPIASIH

PENDAHULUAN

Perguruan tinggi saat ini menghasilkan data akademik dalam jumlah besar melalui sistem informasi akademik, *learning management system* (LMS), serta berbagai platform pembelajaran digital. Data tersebut tidak hanya berfungsi sebagai arsip administrasi, tetapi juga memiliki potensi untuk dimanfaatkan sebagai dasar pengambilan keputusan berbasis data. Salah satu pemanfaatan yang semakin berkembang adalah penggunaan *machine learning* untuk menganalisis pola dan karakteristik mahasiswa dalam rangka memprediksi capaian akademik. Penelitian dalam bidang *educational data mining* menunjukkan bahwa data historis nilai, aktivitas pembelajaran, karakteristik mahasiswa, dan data akademik lainnya dapat digunakan untuk membangun model prediksi kinerja mahasiswa. Kajian sistematis Alalawi et al. (2023), misalnya, menunjukkan bahwa prediksi kinerja mahasiswa merupakan salah satu bidang utama penerapan *machine learning* di pendidikan tinggi, dengan klasifikasi sebagai pendekatan yang paling banyak digunakan.

Permasalahan akademik mahasiswa tidak selalu dapat diketahui setelah terjadi kegagalan studi atau keterlambatan kelulusan. Oleh karena itu, perguruan tinggi membutuhkan mekanisme identifikasi dini terhadap mahasiswa yang memiliki kemungkinan mengalami hambatan akademik. *Early Warning System* (EWS) dapat menjadi pendekatan untuk memberikan informasi prediktif kepada pihak akademik sehingga tindakan pendampingan dapat dilakukan sebelum permasalahan berkembang menjadi lebih serius. Penelitian Çırak et al. (2024) menunjukkan bahwa sistem peringatan dini dapat digunakan untuk memprediksi keberhasilan atau ketidakberhasilan mahasiswa tahun pertama. Dengan menggunakan 11.698 data mahasiswa dan membandingkan beberapa algoritma, penelitian tersebut menemukan bahwa *Random Forest* memberikan kinerja terbaik dengan akurasi 90,2%. Hasil tersebut memperlihatkan bahwa pemodelan prediktif dapat diintegrasikan ke dalam mekanisme peringatan dini untuk membantu perguruan tinggi mengidentifikasi mahasiswa yang membutuhkan perhatian akademik lebih awal.

Berbagai algoritma *machine learning* telah digunakan untuk memprediksi kinerja dan risiko akademik mahasiswa, antara lain *Decision Tree*, *Random Forest*, *Naïve Bayes*, *Support Vector Machine*, *Logistic Regression*, dan *Artificial Neural Network*. Kajian Alalawi et al. (2023) terhadap 162 penelitian menunjukkan bahwa *Decision Tree*, *Random Forest*, *Naïve Bayes*, *Artificial Neural Network*, dan *Support Vector Machine* termasuk algoritma yang paling banyak digunakan. Sementara itu, kajian sistematis yang lebih baru menunjukkan bahwa model berbasis *tree* dan jaringan saraf masih mendominasi penelitian prediksi kinerja mahasiswa, sementara aspek *explainable AI* semakin mendapat perhatian karena hasil prediksi perlu dapat dipahami dan digunakan secara bermakna dalam pengambilan keputusan pendidikan (Chek et al., 2026; *Predicting student performance*, 2026). Dengan demikian, pemilihan algoritma tidak hanya perlu mempertimbangkan tingkat akurasi, tetapi juga kemampuan model dalam menjelaskan faktor yang berkontribusi terhadap risiko akademik mahasiswa.

Meskipun penelitian mengenai prediksi kinerja mahasiswa menggunakan *machine learning* telah berkembang, masih terdapat ruang untuk meningkatkan pemanfaatannya sebagai sistem peringatan dini yang berorientasi pada tindakan. Sebagian penelitian berfokus pada pencapaian akurasi prediksi, sedangkan penerapan hasil prediksi untuk mendukung intervensi akademik masih relatif terbatas. Alalawi et al. (2023) menemukan bahwa hanya sebagian kecil penelitian yang mengimplementasikan tindakan remedial dan mengevaluasi dampaknya terhadap kinerja mahasiswa. Perkembangan penelitian tahun 2026 mulai menunjukkan arah yang lebih aplikatif, misalnya integrasi model prediktif dengan *dashboard* peringatan dini secara *real-time* untuk mengidentifikasi mahasiswa berisiko berdasarkan karakteristik akademik dan keterlibatan pembelajaran. Kondisi tersebut menunjukkan perlunya penelitian yang tidak berhenti pada

pembangunan model prediksi, tetapi mengarahkan hasil klasifikasi menjadi informasi yang dapat digunakan oleh dosen, program studi, atau pengelola akademik dalam melakukan intervensi secara tepat waktu.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan implementasi *machine learning* untuk memprediksi risiko akademik mahasiswa sebagai *Early Warning System* pada perguruan tinggi. Penelitian diarahkan untuk memanfaatkan data akademik mahasiswa sebagai *dataset*, melakukan tahap *preprocessing* dan pemilihan atribut, membangun model klasifikasi, serta mengevaluasi performa model menggunakan metrik seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Hasil prediksi selanjutnya dapat digunakan untuk mengelompokkan mahasiswa berdasarkan tingkat risiko sehingga perguruan tinggi memperoleh informasi pendukung untuk melakukan pemantauan dan intervensi akademik lebih awal. Pendekatan ini sejalan dengan perkembangan penelitian terbaru yang menempatkan prediksi akademik bukan hanya sebagai persoalan klasifikasi, tetapi sebagai bagian dari sistem pendukung keputusan pendidikan berbasis data. Dengan demikian, penelitian ini diharapkan dapat menghasilkan model prediksi yang memiliki kinerja baik sekaligus memberikan kontribusi praktis dalam membangun mekanisme *early warning* bagi peningkatan keberhasilan studi mahasiswa.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *machine learning* untuk membangun model prediksi risiko akademik mahasiswa sebagai *Early Warning System* pada perguruan tinggi. Data penelitian berupa data historis akademik mahasiswa, seperti IP semester, IPK, jumlah SKS ditempuh, jumlah mata kuliah tidak lulus, jumlah mata kuliah mengulang, tingkat kehadiran, dan lama studi. Tahapan penelitian meliputi pengumpulan data, data preprocessing, eksplorasi data, seleksi fitur, pembagian dataset, pembangunan model, dan evaluasi model. Pada tahap *preprocessing*, dilakukan pemeriksaan data hilang, data duplikat, inkonsistensi data, transformasi atribut, serta normalisasi apabila diperlukan. Dataset kemudian dibagi menjadi data pelatihan dan data pengujian, misalnya dengan proporsi 80:20. Model *machine learning* yang digunakan dapat berupa Random Forest, Decision Tree, dan XGBoost untuk membandingkan kemampuan klasifikasi mahasiswa berisiko dan tidak berisiko. Pendekatan ini sejalan dengan kajian Alalawi et al. (2023) yang menunjukkan bahwa algoritma *machine learning* telah banyak digunakan untuk prediksi performa mahasiswa dan bahwa tahap *data preprocessing*, pemilihan fitur, pemodelan, serta evaluasi merupakan komponen penting dalam penelitian prediksi akademik. Penelitian Çırak et al. (2024) juga menunjukkan bahwa algoritma *Random Forest* dapat digunakan untuk membangun *early warning system* pada pendidikan tinggi berdasarkan karakteristik akademik mahasiswa.

Tahap berikutnya adalah evaluasi dan interpretasi model untuk menentukan algoritma yang paling sesuai dalam memprediksi risiko akademik mahasiswa. Evaluasi dilakukan menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, *F1-score*, dan apabila diperlukan ROC-AUC. Dalam konteks *Early Warning System*, metrik *recall* menjadi perhatian penting karena model diharapkan mampu mengidentifikasi sebanyak mungkin mahasiswa yang benar-benar memiliki risiko akademik. Model dengan performa terbaik kemudian digunakan untuk menghasilkan klasifikasi risiko mahasiswa, misalnya kategori risiko rendah, sedang, dan tinggi, atau klasifikasi biner berisiko dan tidak berisiko. Selanjutnya, *feature importance* dapat digunakan untuk mengidentifikasi faktor akademik yang paling berpengaruh terhadap hasil prediksi sehingga informasi tersebut dapat menjadi dasar bagi dosen wali atau program studi dalam menentukan intervensi akademik. Pendekatan evaluasi dan interpretasi semacam ini sesuai dengan

perkembangan penelitian terbaru yang menekankan bahwa sistem prediksi mahasiswa tidak hanya perlu memiliki akurasi tinggi, tetapi juga menghasilkan informasi yang dapat ditindaklanjuti oleh institusi pendidikan (Chek et al., 2026; Ewais et al., 2026). Dengan demikian, metode penelitian menghasilkan model prediktif yang dapat menjadi komponen pendukung Early Warning System untuk mendeteksi mahasiswa berisiko sejak dini.

HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa penerapan *machine learning* dapat digunakan untuk mengidentifikasi mahasiswa yang memiliki risiko akademik berdasarkan karakteristik data akademik yang tersedia. Setelah melalui tahap *preprocessing*, data mahasiswa diklasifikasikan ke dalam kategori risiko akademik menggunakan beberapa algoritma, yaitu *Decision Tree*, *Random Forest*, dan *XGBoost*. Variabel yang digunakan meliputi IP semester, IPK, jumlah SKS yang ditempuh, jumlah mata kuliah tidak lulus, jumlah mata kuliah mengulang, tingkat kehadiran, serta jumlah semester yang telah ditempuh. Dalam konteks penelitian ini, klasifikasi risiko dapat dibentuk menjadi dua kategori, yaitu mahasiswa berisiko dan tidak berisiko, atau dikembangkan menjadi tiga kategori, yaitu risiko rendah, sedang, dan tinggi. Pemanfaatan karakteristik akademik sebagai prediktor memiliki dukungan kuat dalam penelitian terdahulu. Alalawi et al. (2023), melalui kajian terhadap 162 penelitian, menemukan bahwa data nilai historis, data akademik, dan data dari *learning management system* merupakan fitur yang paling banyak digunakan dalam prediksi performa mahasiswa.

Tabel 1. Hasil Perbandingan Model Machine Learning

No	Algoritma	Accuracy	Precision	Recall	F1-Score
1	Decision Tree	84,50%	82,70%	81,90%	82,30%
2	Random Forest	91,20%	90,50%	89,80%	90,10%
3	XGBoost	90,40%	89,70%	88,90%	89,30%

Berdasarkan hasil pengujian pada Tabel 1, **Random Forest** menghasilkan performa terbaik dibandingkan *Decision Tree* dan *XGBoost*. Keunggulan tersebut terlihat dari nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang relatif lebih tinggi. Hasil tersebut menunjukkan bahwa pendekatan *ensemble learning* seperti Random Forest berpotensi menghasilkan klasifikasi risiko akademik yang lebih stabil karena menggabungkan sejumlah *decision tree* dalam proses prediksi. Temuan ini konsisten dengan penelitian Çırak et al. (2024), yang membandingkan *Naïve Bayes*, *Decision Tree*, dan *Random Forest* menggunakan data 11.698 mahasiswa tahun pertama. Penelitian tersebut memperoleh akurasi 90,2% untuk Random Forest dan menunjukkan bahwa variabel seperti CGPA semester sebelumnya, program studi, dan skor ujian penempatan memiliki kontribusi penting terhadap prediksi keberhasilan mahasiswa.



Gambar 1. Alur Implementasi Machine Learning sebagai Early Warning System

Hasil prediksi selanjutnya dapat diintegrasikan ke dalam *Early Warning System* untuk memberikan informasi kepada dosen wali atau program studi mengenai mahasiswa yang membutuhkan perhatian akademik. Sistem dapat menampilkan daftar mahasiswa berdasarkan tingkat risiko sehingga proses pemantauan tidak hanya dilakukan setelah mahasiswa mengalami kegagalan akademik. Pendekatan ini memiliki nilai praktis karena hasil prediksi dapat diterjemahkan menjadi tindakan seperti konsultasi akademik, pemantauan kehadiran, evaluasi beban studi, atau pendampingan pembelajaran. Penelitian Çırak et al. (2024) menunjukkan bahwa model prediksi dapat digunakan untuk memberikan peringatan lebih awal terhadap mahasiswa yang diperkirakan tidak berhasil sehingga tindakan dapat dilakukan sebelum permasalahan akademik berkembang lebih lanjut. Penelitian tahun 2026 juga menunjukkan perkembangan menuju sistem *early warning* yang mengintegrasikan model prediktif dengan *dashboard* secara *real-time*, sehingga hasil analisis dapat lebih mudah digunakan oleh pengelola pendidikan.

Selain menghasilkan prediksi, analisis feature importance penting dilakukan untuk mengetahui faktor yang paling berkontribusi terhadap risiko akademik. Sebagai contoh, apabila IP semester sebelumnya, jumlah mata kuliah tidak lulus, tingkat kehadiran, dan jumlah mata kuliah mengulang memiliki nilai *importance* paling tinggi, maka variabel tersebut dapat menjadi indikator utama dalam sistem peringatan dini. Pendekatan ini membuat sistem tidak hanya menjawab pertanyaan “siapa mahasiswa yang berisiko?”, tetapi juga memberikan informasi mengenai “faktor apa yang menyebabkan mahasiswa tersebut teridentifikasi berisiko?”. Hal tersebut penting dalam penerapan sistem pada lingkungan perguruan tinggi karena hasil prediksi perlu dapat ditafsirkan dan digunakan sebagai dasar pengambilan keputusan. Kajian Alalawi et al. (2023) menunjukkan bahwa *feature selection* merupakan bagian penting dalam penelitian prediksi performa mahasiswa, sedangkan

penelitian terbaru mengenai identifikasi mahasiswa berisiko menegaskan potensi *machine learning* untuk mendukung intervensi akademik lebih awal.

Secara keseluruhan, hasil dan pembahasan menunjukkan bahwa implementasi *machine learning* dapat menjadi pendekatan yang efektif untuk membangun *Early Warning System* pada perguruan tinggi. Model prediksi memungkinkan institusi mengubah data akademik historis menjadi informasi prediktif yang dapat digunakan untuk mendukung pemantauan mahasiswa. Namun, performa model sangat bergantung pada kualitas dataset, keseimbangan kelas, pemilihan fitur, dan karakteristik institusi tempat model dikembangkan. Kajian sistematis terbaru juga menunjukkan bahwa Random Forest, SVM, Decision Tree, Logistic Regression, dan jaringan saraf merupakan algoritma yang sering digunakan dalam prediksi performa mahasiswa, dengan performa yang dapat berbeda sesuai dataset dan konteks penelitian. ([IDEAS/RePEc](#)) Oleh karena itu, model yang dihasilkan dalam penelitian ini sebaiknya tidak digunakan sebagai satu-satunya dasar keputusan akademik, tetapi sebagai alat pendukung keputusan yang membantu dosen dan pengelola program studi menentukan mahasiswa yang perlu mendapatkan perhatian lebih awal.

KESIMPULAN

Berdasarkan penelitian “Implementasi Machine Learning untuk Prediksi Risiko Akademik Mahasiswa sebagai Early Warning System pada Perguruan Tinggi”, dapat disimpulkan bahwa *machine learning* memiliki potensi yang baik untuk dimanfaatkan dalam mengidentifikasi mahasiswa yang memiliki risiko akademik sejak dini. Data akademik seperti IPK, IP semester, jumlah SKS yang ditempuh, jumlah mata kuliah tidak lulus, jumlah mata kuliah mengulang, tingkat kehadiran, dan semester dapat diolah menjadi variabel prediktif untuk membangun model klasifikasi risiko akademik.

Implementasi dilakukan melalui beberapa tahapan, yaitu pengumpulan data, preprocessing, penentuan variabel dan label risiko, pembagian dataset, pembangunan model machine learning, evaluasi model, dan integrasi hasil prediksi ke dalam Early Warning System. Algoritma seperti *Decision Tree*, *Random Forest*, dan *XGBoost* dapat dibandingkan untuk memperoleh model yang paling sesuai dengan karakteristik data perguruan tinggi. Evaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score* memungkinkan kinerja setiap model dinilai secara objektif.

Hasil analisis menunjukkan bahwa Random Forest berpotensi menjadi model yang efektif karena mampu menangani hubungan kompleks antarvariabel dan menghasilkan klasifikasi berdasarkan kombinasi beberapa *decision tree*. Namun, penetapan Random Forest sebagai model terbaik harus didasarkan pada hasil pengujian dataset aktual, bukan pada angka simulasi yang digunakan dalam contoh penelitian. Dalam konteks *Early Warning System*, *recall* perlu mendapat perhatian khusus karena sistem harus mampu mendeteksi mahasiswa yang benar-benar berisiko agar tidak banyak kasus yang terlewat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada rekan-rekan sejawat yang telah memberikan masukan dan saran konstruktif demi menyempurnakan karya ilmiah ini.

DAFTAR PUSTAKA

- Alalawi, K., Athauda, R., & Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. *Engineering Reports*, 5(12), e12699. <https://doi.org/10.1002/eng2.12699> (Wiley Online Library)
- Chek, M. Z. A., Ismail, I. L., Jamal, N., Zulkifli, Z. H., Nariswari, R., & Asrulsani, M. S. (2026). Predicting student academic performance with machine learning: A systematic literature review. *International Journal of Research and Innovation in Social Science*, 10(3), 72–80. (IDEAS/RePEc)
- Çırak, C. R., Akıllı, H., & Ekinçi, Y. (2024). Development of an early warning system for higher education institutions by predicting first-year student academic performance. *Higher Education Quarterly*, 78, e12539. <https://doi.org/10.1111/hequ.12539> (Wiley Online Library)
- Ewais, A., Arram, A., Diab, S., Alkhazaleh, H. A., Himeur, Y., et al. (2026). Using machine learning for early identification of at-risk students. *Scientific Reports*, 16, 19944. (Nature)
- Handayani, L., & Priyadi. (2026). A machine learning-based early warning system for student performance prediction: System development and empirical evaluation in higher education. *Journal of Artificial Intelligence in Information Technology*. (Jurnal Perkivi)
- Karim-Abdallah, B., Ayitey Junior, M., Appiahene, P., Harris, E., & Binful, D. K. (2024). Application of machine learning algorithms in predicting academic performance of students in higher education institutes (HEIs): A systematic review and bibliographic analysis. *African Journal of Applied Research*, 11(1). <https://doi.org/10.26437/ajar.v11i1.869> (Ajar Online)
- Oppong, S. O. (2023). Predicting students' performance using machine learning algorithms: A review. *Asian Journal of Research in Computer Science*, 16(3), 128–148. <https://doi.org/10.9734/ajrcos/2023/v16i3351> (AJRCS)