Journal of Innovative and Creativity
Journal Homepage: https://joecy.org/index.php/joecy
e-mail: journaljoecy@gmail.com

Faculty of Education
University of Pahlawan Tuanku Tambusai
ISSN 2775-771X (Online)
ISSN 2962-570X **(Printed)**
6 (2) 2026, Page: 29083-29097

Literature Review: Penerapan Algoritma Machine Learning pada Sistem Informasi Akademik untuk Prediksi Kelulusan Mahasiswa Tepat Waktu

Anofrizen[1], Syaifullah[2], Niki Rahmah[3], Salsa Abdi Ramadhani[4], Azalia Azahra Sandero[5], Fadhila Salsabila Lubis[6]
[123456]Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Kota Pekanbaru,Tampan, Riau, Indonesia
Corresponding Author: Syaifullah syaifullah@uin-suska.ac.id

Published: May , 2026

ABSTRAK

Kelulusan mahasiswa tepat waktu merupakan salah satu indikator penting dalam menilai kualitas akademik, efektivitas pengelolaan pendidikan tinggi, dan keberhasilan layanan akademik perguruan tinggi. Sistem informasi akademik menyimpan berbagai data historis mahasiswa, seperti indeks prestasi, nilai semester, jumlah kredit, status akademik, program studi, dan riwayat kelulusan yang dapat dimanfaatkan untuk mendukung proses prediksi. Penelitian ini bertujuan untuk mengkaji penerapan algoritma machine learning pada sistem informasi akademik dalam memprediksi kelulusan mahasiswa tepat waktu. Metode yang digunakan adalah literature review dengan menelaah artikel-artikel ilmiah yang relevan mengenai prediksi kelulusan mahasiswa, prediksi waktu studi, klasifikasi kelulusan, optimasi algoritma, seleksi fitur, dan penanganan data tidak seimbang. Hasil kajian menunjukkan bahwa algoritma *Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest, XGBoost, Artificial Neural Network, deep learning, stacking, bagging, dan boosting* banyak digunakan dalam prediksi kelulusan mahasiswa. Teknik praproses seperti data cleaning, transformasi data, normalisasi, SMOTE, random oversampling, feature selection, dan hyperparameter tuning berperan penting dalam meningkatkan performa model. Kajian ini menunjukkan bahwa machine learning berpotensi besar untuk diintegrasikan dengan sistem informasi akademik sebagai alat bantu deteksi dini mahasiswa yang berisiko terlambat lulus.

Kata kunci: Machine Learning, Sistem Informasi Akademik, Prediksi Kelulusan, Kelulusan Tepat Waktu, Literature Review

ABSTRACT

Student completion is an important indicator for assessing education quality, the effectiveness of higher education management, and the success of academic services in higher education. Information systems in educational institutions store various historical data on students, such as grade point average (GPA), semester grades, number of credits, academic status, curriculum, and graduation history, which can be used to support the forecasting process. The objective of this study is to examine the application of machine learning algorithms in educational information systems for predicting student completion. The method used is a literature review, including an examination of relevant scientific articles on student completion prediction, time prediction, graduation classification, algorithm optimization, feature selection, and handling imbalanced data. The results show that Naive Bayes, K-Nearest Neighbors, Support Vector Machines, Decision Tree, Random Forest, XGBoost, Artificial Neural Network, Deep Learning, Stacking, Bagging, *and Boosting algorithms are widely used in student completion prediction. Prepossessing methods such as data cleaning, data transformation, normalization, SMOTE, random resampling, feature selection, and hyper parameter tuning play a significant role in improving model performance. This study demonstrates that machine learning has significant potential for integration with academic information systems as a tool for the early identification of students at risk of late graduation.*

Keywords: Machine Learning, Academic Information Systems, Graduation Forecasting, On-time Completion, Literature Review.

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Atibution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0

PENDAHULUAN

Kelulusan mahasiswa tepat waktu menjadi salah satu indikator penting dalam menilai mutu penyelenggaraan pendidikan tinggi (Haryatmi & Hervianti, 2021). Tingkat kelulusan yang baik dapat mencerminkan keberhasilan perguruan tinggi dalam mengelola proses akademik, layanan pembimbingan,

kurikulum, dan sistem evaluasi mahasiswa (Goncalves et al., 2024). Kesumawati dan Utari (2018) menjelaskan bahwa pola kelulusan mahasiswa dapat diprediksi melalui pemanfaatan data akademik dengan algoritma Naïve Bayes dan Support Vector Machine. Prediksi waktu kelulusan mahasiswa juga menjadi bagian penting dari kajian educational data mining karena berkaitan langsung dengan efektivitas pengambilan keputusan akademik (Mohammad Suhaimi et al., 2019).

Sistem informasi akademik berperan sebagai sumber data utama dalam analisis kelulusan mahasiswa karena menyediakan berbagai data akademik dan administratif, seperti nilai semester, indeks prestasi kumulatif, jumlah satuan kredit semester, status mahasiswa, lama studi, asal sekolah, program studi, serta data kelulusan (Anwar & Napila, 2025). Lagman et al. (2020) menggunakan data akademik mahasiswa untuk meningkatkan akurasi prediksi kelulusan melalui model ensemble. Riwayat akademik mahasiswa terbukti dapat digunakan untuk membandingkan efektivitas algoritma Naïve Bayes dan K-Nearest Neighbor dalam memprediksi kelulusan tepat waktu (Riadi et al., 2024).

Pemanfaatan machine learning pada sistem informasi akademik dapat membantu perguruan tinggi melakukan deteksi dini terhadap mahasiswa yang memiliki risiko terlambat menyelesaikan studi (Pawitra et al., 2024). Algoritma machine learning bekerja dengan mempelajari pola dari data historis, kemudian menghasilkan prediksi terhadap data baru (Sarker, 2021). Fahrudy dan 'Uyun (2022) menunjukkan bahwa Naïve Bayes dapat digunakan untuk klasifikasi kelulusan mahasiswa dengan bantuan random oversampling dan seleksi fitur. Kombinasi algoritma klasifikasi dan feature selection juga mampu meningkatkan performa prediksi kelulusan mahasiswa tepat waktu (Setiadi et al., 2024).

Permasalahan utama dalam prediksi kelulusan mahasiswa adalah variasi karakteristik data akademik dan ketidakseimbangan data (Bakri et al., 2025). Data mahasiswa yang lulus tepat waktu sering kali memiliki jumlah yang tidak seimbang dibandingkan mahasiswa yang terlambat lulus, sehingga model dapat cenderung memprediksi kelas mayoritas (Efendi & Defit, 2025). Yaqin et al. (2022) menggunakan Synthetic Minority Oversampling Technique untuk meningkatkan performa ANN, K-NN, dan SVM pada prediksi kelulusan mahasiswa. Penerapan SMOTE juga dapat meningkatkan akurasi, presisi, dan sensitivitas model SVM pada data kelulusan mahasiswa yang tidak seimbang (Anggrawan et al., 2023).

Pengembangan model prediksi kelulusan tidak hanya bergantung pada pemilihan algoritma, tetapi juga pada teknik praproses dan optimasi model (Yağcı, 2022). Teknik seperti resampling, hyperparameter tuning, seleksi fitur, stacking, bagging, dan boosting banyak digunakan untuk meningkatkan akurasi model prediksi (Herianto et al., 2024). Bakri et al. (2025) membandingkan beberapa model machine learning dengan pendekatan resampling dan hyperparameter tuning dalam prediksi kelulusan tepat waktu. Herianto et al. (2024) juga menggunakan teknik stacking untuk mengoptimalkan performa algoritma machine learning dalam prediksi kelulusan mahasiswa.

Perkembangan penelitian menunjukkan bahwa prediksi kelulusan mahasiswa tidak lagi hanya menggunakan model Tunggal (Okoye et al., 2024). Model ensemble dan deep learning mulai banyak digunakan karena mampu menangkap pola data yang lebih kompleks (Rabelo & Zarate, 2025). Nanglae et al. (2021) mengembangkan bi-level learning framework untuk menentukan pola kelulusan mahasiswa. Nguyen et al. (2025) menunjukkan bahwa model deep learning sinergis dapat meningkatkan akurasi klasifikasi kelulusan mahasiswa.

Kajian tentang prediksi kelulusan tepat waktu penting dilakukan untuk melihat pola penelitian terdahulu secara menyeluruh (Pelima et al., 2024). Literature review dapat membantu mengidentifikasi algoritma yang sering digunakan, jenis data atau atribut yang dimanfaatkan, metode prediksi yang diterapkan, serta gap penelitian terdahulu. (Issah et al., 2023). Law et al. (2024) menyatakan bahwa penelitian graduate on time prediction masih memiliki ruang pengembangan, terutama dalam pemilihan variabel penting dan peningkatan performa model. Machine learning dan performa akademik menjadi komponen penting dalam kajian sistematis prediksi kelulusan mahasiswa (Pelima et al., 2024).

Penelitian ini berjudul "Literature Review: Penerapan Algoritma Machine Learning pada Sistem Informasi Akademik untuk Prediksi Kelulusan Mahasiswa Tepat Waktu". Fokus kajian diarahkan pada artikel yang membahas prediksi kelulusan mahasiswa, klasifikasi kelulusan tepat waktu, prediksi durasi studi, dan pengolahan data akademik menggunakan machine learning. Sadqui et al. (2023) menekankan pentingnya evaluasi beberapa model machine learning untuk memahami algoritma yang sesuai dalam memprediksi timeline kelulusan mahasiswa. Model bagging dan boosting juga menjadi perhatian karena mampu memberikan hasil prediksi yang kompetitif dalam konteks kelulusan mahasiswa (Sitio et al.,

2026).

Tujuan penelitian ini adalah mengidentifikasi penerapan algoritma machine learning pada sistem informasi akademik untuk prediksi kelulusan mahasiswa tepat waktu. Kajian ini juga bertujuan untuk mengetahui jenis data akademik yang digunakan, teknik praproses yang diterapkan, metode optimasi yang digunakan, dan metrik evaluasi yang dipakai dalam penelitian terdahulu. Mohammad Suhaimi et al. (2019) menempatkan prediksi waktu kelulusan mahasiswa sebagai salah satu topik penting dalam kajian machine learning di bidang pendidikan. Pelima et al. (2024) juga menunjukkan bahwa kajian sistematis dapat membantu merangkum perkembangan penelitian terkait prediksi kelulusan mahasiswa berbasis performa akademik dan machine learning.

METODE

Metode yang digunakan dalam penelitian ini adalah literature review dengan pendekatan Systematic Literature Review. Pendekatan ini digunakan untuk mengidentifikasi, menyeleksi, menganalisis, dan merangkum penelitian terdahulu yang relevan dengan penerapan algoritma machine learning dalam prediksi kelulusan mahasiswa tepat waktu. Mohammad Suhaimi et al. (2019) menggunakan pendekatan review untuk memetakan perkembangan algoritma machine learning dalam prediksi waktu kelulusan mahasiswa. Kajian sistematis juga dapat memberikan gambaran mengenai hubungan performa akademik, data pendidikan, dan penerapan machine learning dalam prediksi kelulusan (Pelima et al., 2024).

Proses literature review dalam penelitian ini dilakukan melalui beberapa tahapan, mulai dari penentuan topik, pencarian artikel, seleksi artikel, analisis isi artikel, sampai penyusunan hasil kajian. Tahapan tersebut digunakan agar artikel yang dikaji benar-benar sesuai dengan fokus penelitian, yaitu penerapan algoritma machine learning pada sistem informasi akademik untuk prediksi kelulusan mahasiswa tepat waktu. Alur metode penelitian ditampilkan pada Gambar 1.

Gambar 1. Alur Metode Literature Review



Gambar 1 menunjukkan tahapan literature review yang digunakan dalam penelitian ini. Tahap pertama adalah menentukan topik penelitian yang berfokus pada prediksi kelulusan mahasiswa tepat waktu menggunakan algoritma machine learning. Tahap kedua adalah menentukan kata kunci pencarian yang sesuai dengan topik, seperti machine learning, student graduation prediction, on-time graduation prediction, dan academic information system. Tahap berikutnya adalah menyeleksi artikel berdasarkan judul, abstrak, dan isi artikel untuk memastikan kesesuaiannya dengan kriteria penelitian. Artikel yang lolos seleksi kemudian dikelompokkan berdasarkan algoritma, teknik praproses, feature selection, penanganan data tidak seimbang, serta metrik evaluasi.

Strategi pencarian literatur dilakukan dengan menentukan topik, kata kunci, sumber artikel, dan kriteria seleksi. Kata kunci yang digunakan meliputi "machine learning", "student graduation prediction", "on-time graduation prediction", "graduation time prediction", "academic information system", "student academic performance", dan "educational data mining". Law et al. (2024) menggunakan fokus graduate on time prediction untuk menelaah metode, isu, dan arah penelitian prediksi kelulusan tepat waktu. Artikel yang dikaji diarahkan pada penelitian yang membahas

penggunaan data akademik dan algoritma machine learning untuk memprediksi status kelulusan mahasiswa (Kesumawati & Utari, 2018).

Artikel yang digunakan dalam penelitian ini dipilih berdasarkan kriteria inklusi dan eksklusi. Kriteria tersebut digunakan untuk membatasi artikel agar pembahasan tetap fokus pada prediksi kelulusan mahasiswa tepat waktu dengan algoritma machine learning. Kriteria inklusi dan eksklusi ditampilkan pada Tabel 1.

Tabel 1. Kriteria Inklusi dan Eksklusi Artikel

Kriteria Inklusi	**Kriteria Eksklusi**
Artikel membahas prediksi kelulusan mahasiswa	Artikel hanya membahas kepuasan mahasiswa
Artikel membahas kelulusan tepat waktu atau waktu kelulusan	Artikel hanya membahas loyalitas pengguna sistem
Artikel menggunakan algoritma machine learning	Artikel tidak menggunakan metode machine learning
Artikel menggunakan data akademik mahasiswa	Artikel tidak menggunakan data mahasiswa
Artikel menjelaskan algoritma, dataset, atau hasil evaluasi	Artikel tidak menjelaskan metode dan hasil evaluasi
Artikel relevan dengan sistem informasi akademik atau educational data mining	Artikel hanya membahas e-learning umum tanpa prediksi kelulusan

Tabel 1 menunjukkan batasan artikel yang digunakan dalam penelitian. Artikel yang dipilih harus memiliki hubungan langsung dengan prediksi kelulusan mahasiswa, prediksi waktu studi, atau klasifikasi kelulusan tepat waktu. Artikel yang hanya membahas kepuasan pengguna, loyalitas sistem, atau performa pembelajaran tanpa kaitan dengan kelulusan mahasiswa tidak dimasukkan sebagai sumber utama. Kriteria ini dibuat agar hasil kajian tetap sesuai dengan tujuan penelitian.

Kriteria inklusi dalam penelitian ini mencakup artikel yang membahas prediksi kelulusan mahasiswa, prediksi kelulusan tepat waktu, prediksi waktu studi, atau klasifikasi status kelulusan menggunakan algoritma machine learning. Artikel yang dipilih harus menjelaskan algoritma yang digunakan, dataset, variabel akademik, teknik praproses, atau hasil evaluasi model. Fahrudy dan 'Uyun (2022) memenuhi kriteria karena membahas klasifikasi kelulusan mahasiswa menggunakan Naïve Bayes, random oversampling, information gain, dan forward selection. Penelitian tentang SMOTE juga dimasukkan karena teknik tersebut berkaitan langsung dengan penanganan data tidak seimbang dalam prediksi kelulusan mahasiswa (Yaqin et al., 2022).

Kriteria eksklusi dalam penelitian ini mencakup artikel yang tidak berhubungan langsung dengan prediksi kelulusan mahasiswa tepat waktu. Artikel yang hanya membahas kepuasan mahasiswa, loyalitas pengguna sistem, performa bahasa, evaluasi dosen, atau sistem e-learning umum tidak dijadikan sumber utama. Bakri et al. (2025) termasuk sumber relevan karena membahas optimasi model machine learning untuk prediksi kelulusan tepat waktu. Artikel yang mengkaji algoritma klasifikasi dan feature selection juga digunakan karena berhubungan langsung dengan peningkatan akurasi prediksi kelulusan mahasiswa (Setiadi et al., 2024).

Tahap seleksi literatur dilakukan melalui beberapa langkah, yaitu identifikasi artikel, penyaringan berdasarkan judul dan abstrak, pemeriksaan isi artikel, dan pemilihan artikel akhir. Artikel yang lolos seleksi kemudian dikelompokkan berdasarkan fokus penelitian, algoritma, data, teknik praproses, dan hasil evaluasi. Riadi et al. (2024) dianalisis dalam kelompok penelitian yang membandingkan Naïve Bayes dan K-NN berdasarkan riwayat akademik mahasiswa. Penelitian berbasis ensemble learning dianalisis dalam kelompok tersendiri karena menggunakan kombinasi beberapa model untuk meningkatkan akurasi prediksi (Lagman et al., 2020).

Analisis literatur dilakukan dengan membaca bagian pendahuluan, metode, dataset, algoritma, evaluasi, hasil, dan kesimpulan dari setiap artikel. Artikel kemudian diklasifikasikan berdasarkan jenis algoritma yang digunakan, seperti Naïve Bayes, K-NN, SVM, Decision Tree, Random Forest, XGBoost, ANN, deep learning, stacking, bagging, dan boosting. Herianto et al. (2024) dianalisis dalam kategori optimasi stacking karena penelitian tersebut membahas peningkatan performa model machine learning untuk prediksi kelulusan. Nguyen et al. (2025) dikaji sebagai penelitian yang menunjukkan perkembangan metode dari model klasik menuju deep learning sinergis.

Pemetaan literatur dilakukan untuk mengetahui kecenderungan metode yang digunakan dalam penelitian prediksi kelulusan mahasiswa. Kategori pemetaan mencakup tujuan penelitian, sumber data,

variabel akademik, algoritma, praproses, seleksi fitur, penanganan data tidak seimbang, metrik evaluasi, dan hasil model. Nanglae et al. (2021) mengkaji pola kelulusan mahasiswa melalui kerangka pembelajaran yang menekankan hubungan antara data akademik dan pola kelulusan. Evaluasi beberapa model machine learning juga diperlukan untuk melihat kesesuaian algoritma dalam memprediksi timeline kelulusan mahasiswa (Sadqui et al., 2023).

Evaluasi terhadap artikel dilakukan berdasarkan metrik yang dilaporkan oleh masing-masing penelitian. Metrik yang paling sering ditemukan dalam artikel meliputi accuracy, precision, recall, F1-score, sensitivity, specificity, ROC-AUC, dan confusion matrix. Anggrawan et al. (2023) menggunakan akurasi, presisi, dan sensitivitas untuk menilai peningkatan performa SVM setelah penerapan SMOTE. Setiadi et al. (2024) menggunakan accuracy, precision, recall, dan ROC-AUC dalam membandingkan algoritma klasifikasi dengan teknik feature selection.

Hasil analisis disusun secara deskriptif untuk menunjukkan algoritma yang paling sering digunakan dan teknik yang berpengaruh terhadap peningkatan performa prediksi. Analisis juga diarahkan untuk menemukan tantangan penelitian, seperti data tidak seimbang, keterbatasan jumlah data, pemilihan variabel, dan minimnya integrasi model dengan sistem informasi akademik. Sitio et al. (2026) menunjukkan bahwa pendekatan ensemble seperti Random Forest dan XGBoost dapat dibandingkan untuk menemukan model yang lebih sesuai dalam prediksi kelulusan. Bakri et al. (2025) juga menekankan pentingnya resampling dan hyperparameter tuning dalam meningkatkan prediksi kelulusan tepat waktu.

HASIL DAN PEMBAHASAN

1. Hasil Kajian Literatur

Hasil kajian menunjukkan bahwa penelitian prediksi kelulusan mahasiswa tepat waktu banyak menggunakan data akademik sebagai sumber utama. Data akademik dianggap relevan karena mencerminkan proses belajar mahasiswa sejak awal perkuliahan hingga tahap akhir studi. Kesumawati dan Utari (2018) menggunakan data mahasiswa untuk memprediksi pola tingkat kelulusan dengan Naïve Bayes dan SVM. Riwayat akademik mahasiswa juga digunakan untuk menentukan kemungkinan kelulusan tepat waktu dalam penelitian berbasis Naïve Bayes dan K-NN (Riadi et al., 2024).

Artikel yang dikaji menunjukkan variasi penggunaan algoritma machine learning dalam memprediksi kelulusan mahasiswa. Algoritma yang paling banyak ditemukan meliputi Naïve Bayes, K-NN, SVM, Decision Tree, Random Forest, XGBoost, ANN, deep learning, dan ensemble learning. Mohammad Suhaimi et al. (2019) mencatat bahwa berbagai algoritma machine learning telah digunakan untuk memprediksi waktu kelulusan mahasiswa. Pelima et al. (2024) juga menemukan bahwa performa akademik merupakan salah satu indikator penting dalam penelitian prediksi kelulusan mahasiswa berbasis machine learning.

Beberapa penelitian berfokus pada peningkatan akurasi melalui teknik praproses dan penanganan data tidak seimbang. Data kelulusan sering kali tidak seimbang karena jumlah mahasiswa yang lulus tepat waktu dan tidak tepat waktu dapat berbeda secara signifikan. Yaqin et al. (2022) menggunakan SMOTE untuk memperbaiki distribusi kelas pada data kelulusan mahasiswa. Penerapan SMOTE pada model SVM dapat meningkatkan kemampuan model dalam mengenali kelas minoritas secara lebih baik (Anggrawan et al., 2023).

Penelitian lain menekankan pentingnya feature selection dalam meningkatkan performa model prediksi. Feature selection digunakan untuk memilih variabel yang paling relevan agar model tidak memproses data yang kurang penting. Fahrudy dan 'Uyun (2022) membandingkan information gain dan forward selection dalam klasifikasi kelulusan mahasiswa menggunakan Naïve Bayes. Setiadi et al. (2024) juga menunjukkan bahwa forward selection dan backward elimination dapat meningkatkan performa beberapa algoritma klasifikasi dalam prediksi kelulusan tepat waktu.

Model ensemble menjadi salah satu pendekatan yang banyak dikembangkan dalam penelitian terbaru. Ensemble learning menggabungkan beberapa model untuk memperoleh hasil yang lebih stabil dan akurat. Lagman et al. (2020) menggunakan ensemble model untuk meningkatkan akurasi prediksi kelulusan mahasiswa. Sitio et al. (2026) membandingkan Random Forest sebagai metode bagging dan XGBoost sebagai metode boosting dalam konteks prediksi kelulusan mahasiswa.

Artikel yang terpilih kemudian dirangkum berdasarkan penulis, fokus penelitian, algoritma atau metode yang digunakan, dan hasil utama penelitian. Rangkuman ini bertujuan untuk memperlihatkan kecenderungan penelitian terdahulu dalam memanfaatkan machine learning untuk prediksi kelulusan

mahasiswa tepat waktu. Rangkuman hasil kajian artikel ditampilkan pada Tabel 2.

Tabel 2. Rangkuman Artikel Prediksi Kelulusan Mahasiswa

No	Penulis	Fokus Penelitian	Metode/Algoritma	Hasil Utama
1	Kesumawati & Utari (2018)	Prediksi pola tingkat kelulusan mahasiswa	Naïve Bayes, SVM	SVM memperoleh hasil lebih baik dibandingkan Naïve Bayes
2	Mohammad Suhaimi et al. (2019)	Review prediksi waktu kelulusan mahasiswa	Literature Review	Machine learning banyak digunakan untuk prediksi waktu kelulusan
3	Lagman et al. (2020)	Peningkatan akurasi prediksi kelulusan	Ensemble Model	Ensemble dapat meningkatkan akurasi prediksi kelulusan
4	Nanglae et al. (2021)	Pola kelulusan mahasiswa	Bi-level Learning Framework	Model dapat menentukan pola kelulusan berdasarkan data akademik
5	Fahrudy & 'Uyun (2022)	Klasifikasi kelulusan mahasiswa	Naïve Bayes, Random Oversampling, Information Gain, Forward Selection	Forward selection meningkatkan performa Naïve Bayes
6	Yaqin et al. (2022)	Prediksi awal kelulusan mahasiswa	ANN, K-NN, SVM, SMOTE	SMOTE meningkatkan performa model pada data tidak seimbang
7	Anggrawan et al. (2023)	Peningkatan performa SVM pada data tidak seimbang	SVM, SMOTE	SMOTE meningkatkan akurasi, presisi, dan sensitivitas
8	Sadqui et al. (2023)	Prediksi timeline kelulusan mahasiswa	Beberapa model machine learning	Evaluasi beberapa model diperlukan untuk menentukan algoritma terbaik
9	Setiadi et al. (2024)	Prediksi kelulusan tepat waktu	NB, KNN, DT, Forward Selection, Backward Elimination	KNN dengan forward selection memperoleh performa terbaik
10	Riadi et al. (2024)	Prediksi kelulusan tepat waktu berdasarkan riwayat akademik	Naïve Bayes, K-NN	Naïve Bayes memperoleh akurasi lebih tinggi daripada K-NN
11	Bakri et al. (2025)	Optimasi model prediksi kelulusan tepat waktu	Resampling, Hyperparameter Tuning	Optimasi dapat meningkatkan performa model
12	Nguyen et al. (2025)	Peningkatan akurasi klasifikasi kelulusan	Deep Learning	Model sinergis meningkatkan akurasi klasifikasi
13	Sitio et al. (2026)	Perbandingan bagging dan boosting	Random Forest, XGBoost	XGBoost unggul pada recall dan F1-score

Tabel 2 menunjukkan bahwa penelitian prediksi kelulusan mahasiswa menggunakan berbagai jenis algoritma machine learning. Algoritma klasik seperti Naïve Bayes, K-NN, SVM, dan Decision Tree masih banyak digunakan karena mudah diterapkan pada data akademik. Penelitian terbaru mulai mengarah pada model ensemble, deep learning, resampling, dan hyperparameter tuning untuk meningkatkan performa prediksi. Tabel ini juga memperlihatkan bahwa praproses data dan feature selection menjadi bagian penting dalam peningkatan akurasi model.

2. Algoritma Machine Learning untuk Prediksi Kelulusan

Algoritma machine learning yang digunakan dalam penelitian prediksi kelulusan mahasiswa dapat dikelompokkan berdasarkan jenis pendekatannya. Pengelompokan ini memudahkan pembahasan karena setiap algoritma memiliki karakteristik dan cara kerja yang berbeda. Pengelompokan algoritma

ditampilkan pada Tabel 3.

Tabel 3. Pengelompokan Algoritma Machine Learning

Kelompok Algoritma	Algoritma	Karakteristik Umum
Probabilistik	Naïve Bayes	Menggunakan peluang untuk menentukan kelas kelulusan
Instance-based Learning	K-Nearest Neighbor	Menentukan kelas berdasarkan kedekatan data
Margin-based Learning	Support Vector Machine	Mencari batas pemisah terbaik antar kelas
Tree-based Learning	Decision Tree	Membentuk aturan keputusan dari atribut akademik
Ensemble Learning	Random Forest, XGBoost, Bagging, Boosting, Stacking	Menggabungkan beberapa model untuk meningkatkan performa
Neural Network	ANN, Deep Learning	Mempelajari pola kompleks dan hubungan nonlinier antarvariabel

Tabel 3 menunjukkan bahwa algoritma machine learning dalam prediksi kelulusan mahasiswa dapat dibagi menjadi beberapa kelompok. Naïve Bayes termasuk algoritma probabilistik karena menggunakan perhitungan peluang dalam menentukan kelas. K-NN bekerja berdasarkan kedekatan karakteristik data mahasiswa, sedangkan SVM membangun batas pemisah antar kelas. Model ensemble dan neural network banyak digunakan pada penelitian terbaru karena mampu menangani pola data yang lebih kompleks.

Naïve Bayes merupakan salah satu algoritma yang sering digunakan karena sederhana, cepat, dan mampu bekerja pada data akademik dengan jumlah atribut tertentu. Algoritma ini menggunakan pendekatan probabilistik untuk menentukan kelas kelulusan berdasarkan data historis mahasiswa. Fahrudy dan 'Uyun (2022) menggunakan Naïve Bayes untuk klasifikasi kelulusan mahasiswa dengan bantuan random oversampling dan seleksi fitur. Naïve Bayes juga digunakan dalam penelitian yang membandingkannya dengan K-NN untuk memprediksi kelulusan tepat waktu berdasarkan riwayat akademik (Riadi et al., 2024).

K-Nearest Neighbor digunakan karena mampu mengelompokkan data berdasarkan kedekatan karakteristik antar mahasiswa. Algoritma ini bekerja dengan membandingkan data mahasiswa baru dengan data historis yang memiliki karakteristik serupa. Riadi et al. (2024) membandingkan Naïve Bayes dan K-NN dalam memprediksi kelulusan tepat waktu mahasiswa program pascasarjana. Setiadi et al. (2024) menunjukkan bahwa K-NN yang dikombinasikan dengan forward selection mampu menghasilkan performa tinggi dalam prediksi kelulusan tepat waktu.

Support Vector Machine banyak digunakan karena mampu membangun batas pemisah yang baik antara kelas lulus tepat waktu dan tidak tepat waktu. Algoritma ini sering digunakan pada kasus klasifikasi dengan data akademik yang memiliki atribut numerik dan kategorik. Anggrawan et al. (2023) menerapkan SVM pada data kelulusan mahasiswa yang tidak seimbang dan meningkatkan performanya menggunakan SMOTE. Kesumawati dan Utari (2018) juga menggunakan SVM sebagai salah satu model untuk memprediksi pola tingkat kelulusan mahasiswa.

Decision Tree dan Random Forest banyak digunakan karena mudah dipahami dan dapat memberikan gambaran mengenai faktor yang memengaruhi prediksi. Decision Tree bekerja dengan membentuk aturan keputusan berdasarkan atribut yang tersedia, sedangkan Random Forest menggabungkan banyak pohon keputusan untuk meningkatkan stabilitas model. Setiadi et al. (2024) menggunakan Decision Tree sebagai salah satu algoritma yang dibandingkan dalam prediksi kelulusan tepat waktu. Sitio et al. (2026) menggunakan Random Forest sebagai pendekatan bagging untuk memprediksi kelulusan mahasiswa.

XGBoost menjadi salah satu algoritma boosting yang banyak digunakan dalam penelitian prediksi karena mampu memperbaiki kesalahan model sebelumnya secara bertahap. Algoritma ini efektif dalam menangani data kompleks dan dapat memberikan performa tinggi pada kasus klasifikasi. Sitio et al. (2026) menemukan bahwa XGBoost memiliki recall dan F1-score yang baik dalam memprediksi kelulusan mahasiswa. Bakri et al. (2025) juga menunjukkan bahwa optimasi model, termasuk penggunaan teknik tuning, dapat meningkatkan performa algoritma dalam prediksi kelulusan tepat

waktu.

Artificial Neural Network dan deep learning digunakan untuk menangkap pola nonlinier pada data akademik mahasiswa. Model ini dapat memproses hubungan kompleks antarvariabel, terutama jika dataset memiliki jumlah data dan atribut yang cukup. Yaqin et al. (2022) membandingkan ANN, K-NN, dan SVM dalam prediksi kelulusan mahasiswa dengan bantuan SMOTE. Nguyen et al. (2025) mengembangkan model deep learning sinergis untuk meningkatkan akurasi klasifikasi kelulusan mahasiswa.

3. Data dan Variabel yang Digunakan

Data akademik menjadi komponen utama dalam penelitian prediksi kelulusan mahasiswa tepat waktu. Variabel yang sering digunakan meliputi nilai semester, IPK, jumlah SKS, status mahasiswa, program studi, jenis kelamin, asal sekolah, tahun angkatan, dan lama studi. Kesumawati dan Utari (2018) menggunakan beberapa atribut mahasiswa untuk memprediksi pola kelulusan dengan Naïve Bayes dan SVM. Riwayat akademik seperti IPK sarjana, IPK magister, TOEFL, dan lama studi juga digunakan dalam prediksi kelulusan tepat waktu (Riadi et al., 2024).

Data akademik menjadi dasar utama dalam membangun model prediksi kelulusan mahasiswa. Beberapa variabel sering digunakan karena dianggap mampu merepresentasikan proses dan capaian akademik mahasiswa. Variabel yang sering digunakan dalam penelitian prediksi kelulusan ditampilkan pada Tabel 4.

Tabel 4. Variabel yang Digunakan dalam Prediksi Kelulusan Mahasiswa

No	Variabel	Fungsi dalam Prediksi
1	IPK	Mengukur capaian akademik keseluruhan mahasiswa
2	Nilai semester	Menilai perkembangan akademik mahasiswa setiap semester
3	Jumlah SKS	Mengetahui progres penyelesaian beban studi
4	Lama studi	Menentukan status tepat waktu atau terlambat
5	Program studi	Melihat perbedaan pola akademik antar program
6	Jenis kelamin	Digunakan sebagai variabel demografis pendukung
7	Asal sekolah/universitas	Melihat latar belakang pendidikan mahasiswa
8	Tahun angkatan	Mengelompokkan mahasiswa berdasarkan periode masuk
9	Status kelulusan	Menjadi label kelas dalam model klasifikasi
10	Nilai TOEFL atau tes akademik	Digunakan sebagai indikator kemampuan tambahan

Tabel 4 menunjukkan bahwa IPK, nilai semester, jumlah SKS, dan lama studi merupakan variabel yang paling dekat dengan prediksi kelulusan mahasiswa. Variabel tersebut berhubungan langsung dengan capaian akademik dan progres studi mahasiswa. Variabel demografis seperti jenis kelamin, asal sekolah, dan tahun angkatan digunakan sebagai pendukung untuk memperkuat model. Status kelulusan berfungsi sebagai label kelas yang digunakan untuk membedakan mahasiswa lulus tepat waktu dan tidak tepat waktu.

IPK atau nilai akademik menjadi variabel yang paling banyak digunakan karena mencerminkan kemampuan akademik mahasiswa. Nilai semester awal sering dianggap penting karena dapat menunjukkan potensi keberhasilan studi sejak dini. Setiadi et al. (2024) menggunakan semester achievement index dan atribut akademik lain untuk memprediksi kelulusan tepat waktu. Law et al. (2024) juga menyebut bahwa GPA dan faktor akademik menjadi variabel yang sering muncul dalam penelitian graduate on time prediction.

Jumlah SKS dan capaian kredit juga berperan penting dalam prediksi kelulusan mahasiswa. Mahasiswa yang memiliki capaian SKS sesuai target lebih berpeluang menyelesaikan studi tepat waktu dibandingkan mahasiswa yang tertinggal. Lagman et al. (2020) menggunakan data akademik dan variabel terkait studi untuk membangun model prediksi kelulusan. Nanglae et al. (2021) menekankan bahwa pola kelulusan dapat dipelajari melalui kerangka pembelajaran yang memanfaatkan data akademik mahasiswa.

Variabel demografis juga digunakan dalam beberapa penelitian sebagai pendukung prediksi. Jenis kelamin, usia, asal sekolah, status tempat tinggal, dan sumber pembiayaan dapat memberikan informasi tambahan mengenai karakteristik mahasiswa. Yaqin et al. (2022) menggunakan gender, usia, dan GPA semester sebagai variabel dalam prediksi kelulusan mahasiswa. Setiadi et al. (2024) juga memasukkan data demografis dan finansial untuk melihat faktor yang berpengaruh terhadap kelulusan tepat waktu.

Data dari sistem informasi akademik memiliki potensi besar untuk dimanfaatkan sebagai basis sistem prediksi. Sistem informasi akademik menyimpan data yang terus diperbarui selama mahasiswa

menjalani perkuliahan, sehingga data tersebut dapat digunakan untuk analisis berkelanjutan. Pelima et al. (2024) menegaskan bahwa academic performance merupakan komponen utama dalam prediksi kelulusan mahasiswa berbasis machine learning. Sadqui et al. (2023) juga menggunakan data akademik untuk mengevaluasi beberapa model machine learning dalam memprediksi timeline kelulusan.

4. Teknik Praproses Data

Praproses data merupakan tahap penting sebelum membangun model machine learning. Data akademik sering memiliki nilai kosong, format yang tidak seragam, kategori yang perlu dikodekan, dan skala nilai yang berbeda. Fahrudy dan 'Uyun (2022) melakukan data cleaning, random oversampling, feature selection, dan transformasi data sebelum klasifikasi kelulusan mahasiswa. Setiadi et al. (2024) juga melakukan data cleaning, transformation, random oversampling, dan min-max scaling dalam proses prediksi kelulusan tepat waktu.

Praproses dan optimasi model dilakukan untuk meningkatkan kualitas data serta performa algoritma machine learning. Teknik ini penting karena data akademik sering memiliki nilai kosong, data tidak seimbang, format berbeda, dan fitur yang tidak semuanya relevan. Teknik praproses dan optimasi yang sering digunakan ditampilkan pada Tabel 5.

Tabel 5. Teknik Praproses dan Optimasi Model

Teknik	Tujuan
Data cleaning	Menghapus atau memperbaiki data kosong, duplikat, dan tidak konsisten
Data transformation	Mengubah data kategorik menjadi bentuk numerik
Normalisasi	Menyamakan skala nilai antarvariabel
Label encoding	Mengubah label kategori menjadi angka
One-hot encoding	Mengubah kategori menjadi variabel biner
Random oversampling	Menambah data kelas minoritas secara acak
SMOTE	Membuat data sintetis untuk menyeimbangkan kelas
Feature selection	Memilih atribut yang paling relevan
Hyperparameter tuning	Mencari parameter terbaik pada algoritma
Ensemble learning	Menggabungkan beberapa model untuk meningkatkan performa

Tabel 5 menunjukkan bahwa praproses data tidak hanya dilakukan untuk membersihkan data, tetapi juga untuk membuat data lebih sesuai dengan kebutuhan algoritma. Random oversampling dan SMOTE digunakan untuk menangani data tidak seimbang antara kelas lulus tepat waktu dan tidak tepat waktu. Feature selection membantu mengurangi fitur yang tidak relevan agar model lebih fokus pada atribut penting. Hyperparameter tuning dan ensemble learning digunakan untuk meningkatkan performa model setelah data siap diproses.

Data cleaning dilakukan untuk memperbaiki data yang hilang, duplikat, tidak konsisten, atau tidak sesuai format. Kualitas data sangat menentukan kualitas model karena data yang buruk dapat menghasilkan prediksi yang tidak akurat. Riadi et al. (2024) melakukan pembersihan data sebelum menerapkan Naïve Bayes dan K-NN pada prediksi kelulusan tepat waktu. Anggrawan et al. (2023) juga melakukan praproses sebelum menerapkan SVM dan SMOTE pada data kelulusan mahasiswa.

Transformasi data dilakukan agar data kategorik dapat diproses oleh algoritma machine learning. Teknik yang sering digunakan meliputi label encoding, one-hot encoding, dan konversi atribut kategori menjadi nilai numerik. Riadi et al. (2024) menggunakan transformasi data untuk menyesuaikan atribut akademik dengan kebutuhan algoritma Naïve Bayes dan K-NN. Setiadi et al. (2024) menggunakan transformasi data agar label kelulusan tepat waktu dan tidak tepat waktu dapat diproses dalam model klasifikasi.

Normalisasi data digunakan untuk menyamakan skala antarvariabel numerik. Nilai IPK, jumlah SKS, usia, dan atribut lain dapat memiliki rentang berbeda, sehingga normalisasi membantu algoritma bekerja lebih stabil. Fahrudy dan 'Uyun (2022) menggunakan min-max normalization dalam proses transformasi data sebelum klasifikasi. Setiadi et al. (2024) juga menerapkan min-max scaling untuk menstandarkan nilai fitur sebelum proses klasifikasi.

Penanganan data tidak seimbang menjadi bagian penting dalam praproses prediksi kelulusan mahasiswa. Ketidakseimbangan kelas dapat menyebabkan model menghasilkan prediksi yang bias terhadap kelas mayoritas. Yaqin et al. (2022) menggunakan SMOTE untuk menyeimbangkan data sebelum membandingkan ANN, K-NN, dan SVM. Anggrawan et al. (2023) membuktikan bahwa SMOTE dapat meningkatkan performa SVM pada data kelulusan mahasiswa yang tidak seimbang.

5. Feature Selection dan Optimasi Model

Feature selection digunakan untuk memilih atribut yang paling relevan terhadap prediksi kelulusan mahasiswa. Pemilihan fitur yang tepat dapat mengurangi kompleksitas model, mempercepat proses pelatihan, dan meningkatkan akurasi. Fahrudy dan 'Uyun (2022) membandingkan information gain dan forward selection untuk menentukan fitur terbaik dalam klasifikasi kelulusan mahasiswa. Setiadi et al. (2024) membandingkan forward selection dan backward elimination dalam prediksi kelulusan tepat waktu.

Information gain digunakan untuk mengukur seberapa besar kontribusi suatu atribut terhadap pemisahan kelas. Atribut dengan nilai informasi yang tinggi dianggap lebih relevan dalam proses prediksi. Fahrudy dan 'Uyun (2022) menemukan bahwa information gain dapat meningkatkan performa Naïve Bayes dalam klasifikasi kelulusan mahasiswa. Pemilihan fitur akademik yang relevan juga dapat membantu model memfokuskan proses prediksi pada atribut yang paling berpengaruh (Law et al., 2024).

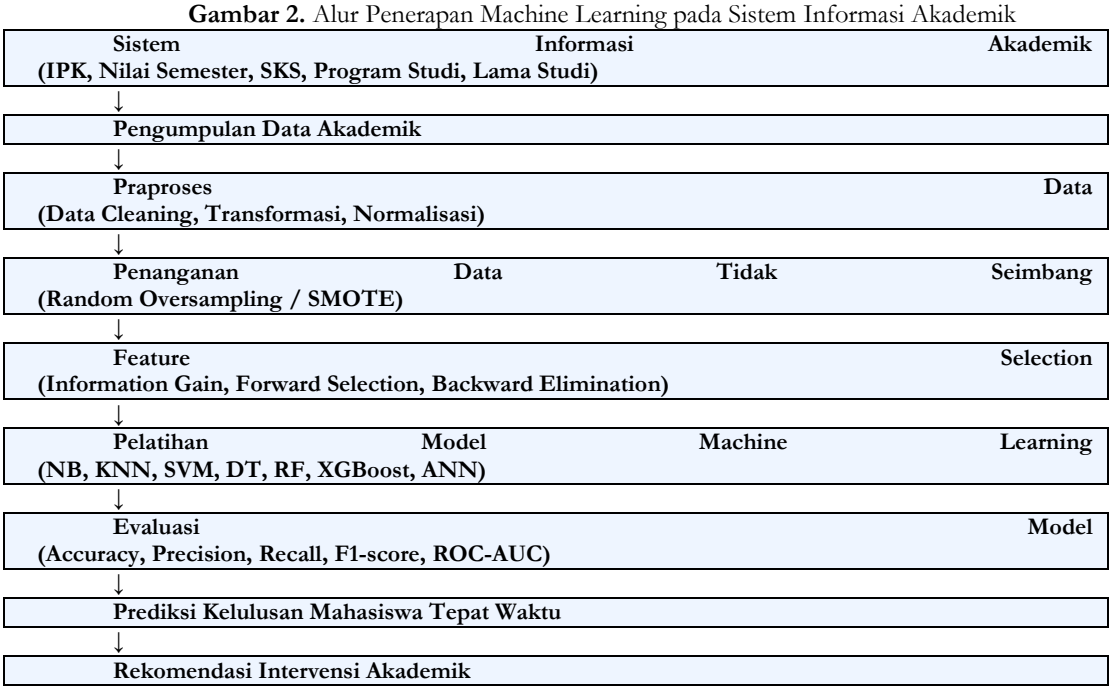
Forward selection bekerja dengan menambahkan fitur secara bertahap berdasarkan kontribusi terhadap performa model. Teknik ini dapat membantu menemukan kombinasi fitur terbaik tanpa menggunakan seluruh atribut dalam dataset. Setiadi et al. (2024) menunjukkan bahwa K-NN dengan forward selection mencapai performa yang baik dalam prediksi kelulusan tepat waktu. Fahrudy dan 'Uyun (2022) juga menemukan bahwa forward selection memberikan performa terbaik pada Naïve Bayes dibandingkan kondisi tanpa seleksi fitur.

Optimasi model dilakukan untuk meningkatkan performa algoritma machine learning. Teknik optimasi dapat berupa hyperparameter tuning, resampling, stacking, bagging, boosting, dan kombinasi beberapa metode. Bakri et al. (2025) membahas optimasi model machine learning untuk prediksi kelulusan tepat waktu melalui resampling dan hyperparameter tuning. Herianto et al. (2024) menggunakan stacking untuk meningkatkan performa algoritma machine learning dalam prediksi kelulusan mahasiswa.

Ensemble learning digunakan untuk menggabungkan beberapa model agar prediksi lebih stabil. Pendekatan ini dapat mengurangi kelemahan model tunggal dan meningkatkan akurasi pada data yang kompleks. Lagman et al. (2020) menunjukkan bahwa ensemble model dapat meningkatkan akurasi prediksi kelulusan mahasiswa. Sitio et al. (2026) juga membandingkan bagging dan boosting untuk melihat pendekatan yang lebih efektif dalam prediksi kelulusan mahasiswa.

Penerapan machine learning dalam prediksi kelulusan mahasiswa dapat digambarkan sebagai alur pemanfaatan data dari sistem informasi akademik sampai menghasilkan keputusan prediktif. Data akademik mahasiswa diproses melalui tahap praproses, pelatihan model, evaluasi, dan hasil prediksi. Hubungan antara sistem informasi akademik dan model machine learning ditampilkan pada Gambar 2.

Gambar 2. Alur Penerapan Machine Learning pada Sistem Informasi Akademik



Gambar 2 menunjukkan bahwa sistem informasi akademik menjadi sumber utama data dalam prediksi kelulusan mahasiswa tepat waktu. Data yang diperoleh dari sistem tersebut perlu melalui tahap praproses agar dapat digunakan oleh algoritma machine learning. Setelah data dibersihkan dan diseimbangkan, model dilatih menggunakan algoritma tertentu dan dievaluasi dengan metrik yang sesuai. Hasil prediksi dapat digunakan sebagai dasar bagi perguruan tinggi untuk memberikan intervensi akademik kepada mahasiswa yang berisiko terlambat lulus.

6. Evaluasi Model

Evaluasi model dilakukan untuk mengetahui seberapa baik algoritma machine learning dalam memprediksi kelulusan mahasiswa tepat waktu. Metrik yang umum digunakan adalah accuracy, precision, recall, F1-score, sensitivity, specificity, ROC-AUC, dan confusion matrix. Anggrawan et al. (2023) menggunakan akurasi, presisi, dan sensitivitas untuk mengukur peningkatan performa SVM setelah penerapan SMOTE. Setiadi et al. (2024) menggunakan accuracy, precision, recall, dan ROC-AUC untuk membandingkan beberapa algoritma klasifikasi.

Evaluasi model digunakan untuk mengetahui kualitas prediksi yang dihasilkan oleh algoritma machine learning. Setiap metrik memiliki fungsi yang berbeda, sehingga penggunaan beberapa metrik dapat memberikan gambaran performa model yang lebih lengkap. Metrik evaluasi yang sering digunakan ditampilkan pada Tabel 6.

Tabel 6. Metrik Evaluasi Model Prediksi Kelulusan

Metrik	Fungsi
Accuracy	Mengukur jumlah prediksi benar dari seluruh data
Precision	Mengukur ketepatan model dalam memprediksi kelas positif
Recall	Mengukur kemampuan model mengenali seluruh data kelas positif
F1-score	Mengukur keseimbangan antara precision dan recall
Sensitivity	Mengukur kemampuan model mendeteksi kelas tertentu
Specificity	Mengukur kemampuan model mengenali kelas negatif
ROC-AUC	Mengukur kemampuan model membedakan dua kelas
Confusion Matrix	Menampilkan rincian TP, TN, FP, dan FN

Tabel 6 menunjukkan bahwa evaluasi model tidak cukup hanya menggunakan accuracy, terutama jika data yang digunakan tidak seimbang. Precision dan recall penting untuk melihat ketepatan serta kemampuan model dalam mengenali mahasiswa yang benar-benar berisiko. F1-score digunakan ketika peneliti ingin melihat keseimbangan antara precision dan recall. Confusion matrix memberikan gambaran lebih rinci mengenai kesalahan dan kebenaran prediksi model.

Accuracy digunakan untuk mengukur persentase prediksi yang benar dari keseluruhan data. Metrik ini banyak digunakan karena mudah dipahami, tetapi kurang cukup jika data tidak seimbang. Yaqin et al. (2022) membandingkan akurasi ANN, K-NN, dan SVM sebelum dan sesudah penerapan SMOTE. Bakri et al. (2025) juga menggunakan akurasi sebagai salah satu metrik utama dalam mengevaluasi model prediksi kelulusan tepat waktu.

Precision digunakan untuk mengukur ketepatan prediksi positif yang diberikan oleh model. Recall digunakan untuk mengukur kemampuan model dalam mengenali seluruh data positif yang seharusnya terdeteksi. Sitio et al. (2026) menekankan pentingnya recall dalam konteks sistem peringatan dini akademik karena kesalahan dalam mendeteksi mahasiswa berisiko dapat berdampak pada keterlambatan intervensi. Anggrawan et al. (2023) juga menunjukkan bahwa sensitivitas meningkat setelah penerapan SMOTE pada SVM.

F1-score digunakan untuk melihat keseimbangan antara precision dan recall. Metrik ini penting ketika dataset tidak seimbang karena accuracy saja dapat menyesatkan. Sitio et al. (2026) menggunakan F1-score untuk membandingkan performa Random Forest dan XGBoost dalam prediksi kelulusan mahasiswa. Setiadi et al. (2024) menggunakan ROC-AUC untuk melihat kemampuan model dalam membedakan kelas kelulusan tepat waktu dan tidak tepat waktu.

Confusion matrix banyak digunakan untuk menganalisis hasil klasifikasi secara lebih rinci. Matriks ini menunjukkan jumlah true positive, true negative, false positive, dan false negative yang dihasilkan model. Riadi et al. (2024) menggunakan confusion matrix untuk mengevaluasi performa Naïve Bayes dan K-NN. Sadqui et al. (2023) juga mengevaluasi model machine learning untuk melihat efektivitas

prediksi timeline kelulusan mahasiswa.

7. Perbandingan Hasil Penelitian

Hasil kajian menunjukkan bahwa tidak ada satu algoritma yang selalu unggul dalam semua kondisi dataset. Performa algoritma sangat dipengaruhi oleh karakteristik data, jumlah fitur, keseimbangan kelas, teknik praproses, dan metode evaluasi. Setiadi et al. (2024) menunjukkan bahwa K-NN dengan forward selection memiliki performa tinggi pada dataset yang digunakan. Riadi et al. (2024) menemukan bahwa Naïve Bayes memiliki akurasi lebih tinggi dibandingkan K-NN dalam prediksi kelulusan tepat waktu berbasis riwayat akademik.

Naïve Bayes cukup sering digunakan karena sederhana dan efisien, tetapi performanya dapat dipengaruhi oleh asumsi independensi antarfitur. K-NN dapat bekerja baik jika fitur yang digunakan relevan dan data telah dinormalisasi. Fahrudy dan 'Uyun (2022) menunjukkan bahwa Naïve Bayes dapat meningkat performanya setelah random oversampling dan feature selection diterapkan. Setiadi et al. (2024) memperlihatkan bahwa K-NN sangat terbantu oleh forward selection dalam meningkatkan accuracy dan ROC-AUC.

SVM memiliki kemampuan yang baik dalam klasifikasi, terutama ketika data telah diproses dengan teknik penanganan ketidakseimbangan. Anggrawan et al. (2023) menunjukkan bahwa SMOTE dapat meningkatkan performa SVM pada data kelulusan mahasiswa yang tidak seimbang. Yaqin et al. (2022) juga membandingkan SVM dengan ANN dan K-NN dalam prediksi kelulusan mahasiswa setelah penerapan SMOTE. Hasil tersebut menunjukkan bahwa SVM dapat menjadi pilihan kuat jika praproses data dilakukan dengan tepat.

Model ensemble dan deep learning menunjukkan potensi yang besar dalam pengembangan sistem prediksi kelulusan. Lagman et al. (2020) menggunakan ensemble model untuk meningkatkan akurasi prediksi kelulusan mahasiswa. Herianto et al. (2024) menggunakan stacking untuk mengoptimalkan performa model machine learning. Nguyen et al. (2025) menunjukkan bahwa model deep learning sinergis dapat meningkatkan akurasi klasifikasi kelulusan mahasiswa.

Penelitian review menunjukkan bahwa graduate on time prediction masih memiliki tantangan dan peluang pengembangan. Tantangan tersebut meliputi pemilihan fitur yang tepat, keterbatasan dataset, ketidakseimbangan data, variasi metode evaluasi, dan integrasi model ke dalam sistem informasi akademik. Law et al. (2024) menekankan perlunya pengembangan penelitian yang mempertimbangkan variabel tambahan dan peningkatan performa model. Pelima et al. (2024) juga menunjukkan bahwa sistem prediksi kelulusan berbasis machine learning masih memerlukan kajian sistematis untuk menyusun arah penelitian berikutnya.

8. Gap Penelitian

Gap pertama yang ditemukan adalah belum semua penelitian menggunakan data yang benar-benar berasal dari sistem informasi akademik secara terintegrasi. Banyak penelitian menggunakan data akademik, tetapi belum selalu menjelaskan bagaimana data tersebut dapat dihubungkan langsung dengan sistem informasi akademik untuk mendukung pengambilan keputusan. Riadi et al. (2024) menggunakan riwayat akademik mahasiswa, sedangkan Setiadi et al. (2024) menggunakan data akademik dari sistem universitas. Integrasi model prediksi ke dalam sistem informasi akademik masih menjadi ruang pengembangan penting (Pelima et al., 2024).

Gap kedua berkaitan dengan keterbatasan variabel yang digunakan dalam model prediksi. Beberapa penelitian hanya menggunakan IPK atau nilai semester, sedangkan faktor lain seperti jumlah SKS, status keuangan, asal sekolah, keaktifan mahasiswa, dan riwayat bimbingan belum selalu dimanfaatkan. Law et al. (2024) menunjukkan bahwa pemilihan variabel penting masih menjadi isu utama dalam graduate on time prediction. Setiadi et al. (2024) menemukan bahwa data akademik, finansial, dan demografis dapat berkontribusi terhadap prediksi kelulusan tepat waktu.

Gap ketiga berkaitan dengan penggunaan metrik evaluasi yang belum seragam. Beberapa penelitian hanya melaporkan accuracy, sedangkan data tidak seimbang membutuhkan metrik lain seperti recall, precision, F1-score, dan ROC-AUC. Anggrawan et al. (2023) menggunakan sensitivitas untuk melihat kemampuan model mengenali kelas tertentu setelah SMOTE. Sitio et al. (2026) menekankan pentingnya recall dan F1-score karena kesalahan deteksi mahasiswa berisiko dapat berdampak pada keterlambatan intervensi akademik.

Gap keempat berkaitan dengan optimasi model yang belum selalu dilakukan secara menyeluruh.

Beberapa penelitian hanya membandingkan algoritma tanpa menerapkan feature selection, resampling, hyperparameter tuning, atau ensemble learning. Bakri et al. (2025) menunjukkan bahwa resampling dan hyperparameter tuning dapat membantu meningkatkan performa model prediksi kelulusan tepat waktu. Herianto et al. (2024) juga membuktikan bahwa stacking dapat digunakan untuk mengoptimalkan algoritma machine learning dalam prediksi kelulusan mahasiswa.

Gap kelima terletak pada penerapan sistem prediksi sebagai alat bantu kebijakan akademik. Banyak penelitian masih berhenti pada tahap pengujian algoritma dan belum mengarah pada implementasi sistem peringatan dini di perguruan tinggi. Sitio et al. (2026) mengaitkan model prediksi dengan kebutuhan Academic Early Warning System. Prediksi kelulusan tepat waktu seharusnya tidak hanya menghasilkan nilai akurasi, tetapi juga dapat membantu perguruan tinggi melakukan intervensi akademik kepada mahasiswa berisiko (Lagman et al., 2020).

Berdasarkan hasil kajian, masih terdapat beberapa gap penelitian yang dapat menjadi arah pengembangan penelitian berikutnya. Gap tersebut berkaitan dengan dataset, metode evaluasi, integrasi sistem, dan optimasi model. Ringkasan gap penelitian ditampilkan pada Tabel 7.

Tabel 7. Gap Penelitian Prediksi Kelulusan Mahasiswa Tepat Waktu

No	Gap Penelitian	Peluang Pengembangan
1	Dataset masih terbatas pada satu institusi	Menggunakan data lintas program studi atau lintas perguruan tinggi
2	Variabel akademik belum lengkap	Menambahkan data bimbingan, aktivitas akademik, dan status keuangan
3	Evaluasi model sering hanya memakai accuracy	Menggunakan precision, recall, F1-score, dan ROC-AUC
4	Integrasi dengan sistem informasi akademik belum banyak dibahas	Mengembangkan sistem prediksi berbasis dashboard akademik
5	Ketidakseimbangan data belum selalu ditangani	Menggunakan SMOTE, random oversampling, atau metode resampling lain
6	Optimasi model belum selalu dilakukan	Menerapkan hyperparameter tuning dan ensemble learning
7	Model prediksi belum selalu digunakan untuk intervensi	Menghubungkan hasil prediksi dengan sistem peringatan dini akademik

Tabel 7 menunjukkan bahwa penelitian prediksi kelulusan mahasiswa tepat waktu masih memiliki banyak ruang pengembangan. Penggunaan dataset yang lebih luas dapat membantu menghasilkan model yang lebih umum dan stabil. Pemilihan variabel yang lebih lengkap dapat meningkatkan kualitas prediksi karena faktor keterlambatan studi tidak hanya dipengaruhi oleh IPK. Integrasi model dengan sistem informasi akademik menjadi peluang penting agar hasil prediksi dapat digunakan langsung dalam pengambilan keputusan akademik.

KESIMPULAN

Penerapan algoritma machine learning pada sistem informasi akademik memiliki potensi besar dalam membantu prediksi kelulusan mahasiswa tepat waktu. Data akademik seperti IPK, nilai semester, jumlah SKS, lama studi, asal pendidikan, status mahasiswa, dan riwayat akademik dapat dimanfaatkan untuk membangun model prediktif. Algoritma yang banyak digunakan dalam penelitian terdahulu meliputi Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest, XGBoost, Artificial Neural Network, deep learning, stacking, bagging, dan boosting. Teknik praproses data memiliki peran penting dalam meningkatkan kualitas model prediksi. Proses data cleaning, transformasi data, normalisasi, label encoding, one-hot encoding, random oversampling, dan SMOTE dapat membantu memperbaiki kualitas data akademik sebelum digunakan dalam pelatihan model. Feature selection seperti information gain, forward selection, dan backward elimination juga dapat mengurangi fitur yang tidak relevan serta meningkatkan performa model. Optimasi seperti hyperparameter tuning, stacking, bagging, dan boosting dapat memperkuat kemampuan model dalam memprediksi kelulusan mahasiswa. Hasil literature review menunjukkan bahwa performa algoritma sangat dipengaruhi oleh karakteristik dataset, keseimbangan kelas, variabel yang digunakan, serta teknik evaluasi. Accuracy sering digunakan sebagai metrik utama, tetapi precision, recall, F1-score, sensitivity, specificity, ROC-AUC, dan confusion matrix lebih tepat digunakan untuk memperoleh penilaian model yang lebih menyeluruh. Prediksi kelulusan mahasiswa tepat waktu tidak hanya penting untuk kepentingan akademik, tetapi juga dapat menjadi dasar pengembangan sistem peringatan dini bagi

perguruan tinggi. Model prediksi yang terintegrasi dengan sistem informasi akademik dapat membantu pengelola pendidikan mengambil tindakan lebih cepat terhadap mahasiswa yang berisiko terlambat lulus.

REFERENSI

Alade, M. S., Adejumo, S. O., Ninan, O. D., Adeniyi, A. E., Ogbuju, E., Aroba, O. J., Ramchander, M., & Adeliyi, T. T. (2025) Prediction of Students' Feedback on Faculty Performance using Stacking Ensemble Method: Machine Leaarning Algorithm. *Edelweiss Applied Science and Technology, 9*(10), 1149-1180.

Alfahid, A. (2024). Algorithmic Prediction of Students On-Time Graduation from the University. *TEM Journal,13*(1), 692-698.

Anggrawan, A., Hairani, H., & Satria, C. (2023). Improving SVM classification performance on unbalanced student graduation time data using SMOTE. International Journal of Information and Education Technology, 13(2), 289–295. https://doi.org/10.18178/ijiet.2023.13.2.1806

Anwar, A. N., & Napila, A. (2025). Naïve Bayes Algorithm Analysis For Student Graduation Timeliness Prediction. *Bit-Tech (Binary Digital - Technology*, *8*(1), 11. https://doi.org/10.32877/bt.v8i1.2825

Arafiyah, R., Santoso, H. B., Hasibuan, Z. A. (2021). Monitoring Learners' Performance by Modeling Learning Progress using Machine Learning. *Journal of Engineering Science and Technology*, 30-39.

Bakri, R., Alam, S., Astuti, N. P., & Bakhtiar, M. I. (2025). Optimizing Machine Learning Models for Graduation on Time Prediction : A Comparative Study with Resampling and Hyperparameter Tuning. *JOIN (Jurnal Online Informatika)*, *10*(2), 270–285. https://doi.org/10.15575/join.v10i2.1590

Efendi, A., & Defit, S. (2025). Addressing Class Imbalance in Machine Learning for Predicting On-Time Student Graduation at The Islamic University of Riau. *JURNAL TEKNIK INFORMATIKA*, *18*(2), 226–235.

Fahrudy, D., & 'Uyun, S. (2022). Classification of student graduation using Naïve Bayes by comparing between random oversampling and feature selections of information gain and forward selection. JOIV: International Journal on Informatics Visualization, 6(4), 798–808. https://doi.org/10.30630/joiv.6.4.982

Goncalves, G. S., Serra, F. A. R., Storopoli, J. E., Scafuto, I. C., & Rafael, D. N. (2024). Undergraduate Student Retention Activities : Challenges and Research Agenda. *SAGE Open*, (September), 1–21. https://doi.org/10.1177/21582440241249334

Haryatmi, E., & Hervianti, S. P. (2021). Penerapan Algoritma Support Vector Machine Untuk Model Preediksi Kelulusan Mahasiswa Tepat Waktu. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *1*(10), 386–392.

Herianto, Kurniawan, B., Hartomi, Z. H., Irawan, Y., & Anam, M. K. (2024). Machine Learning Algorithm Optimization using Stacking Technique for Graduation Prediction. *Journal of Applied Data Sciences*, *5*(3), 1272–1285.

Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, *7*(October 2022), 100204. https://doi.org/10.1016/j.dajour.2023.100204

Kesumawati, A., & Utari, D. T. (2018). Predicting patterns of student graduation rates using Naïve Bayes classifier and support vector machine. AIP Conference Proceedings, 2021, Article 060005. https://doi.org/10.1063/1.5062769

Lagman, A. C., Alfonso, L. P., Goh, M. L. I., Lalata, J. P., Magcuyao, J. P. H., & Vicente, H. N. (2020). Classification algorithm accuracy improvement for student graduation prediction using ensemble model. International Journal of Information and Education Technology, 10(10), 723–727. https://doi.org/10.18178/ijiet.2020.10.10.1449

Law, T.-J., Ting, C.-Y., Ng, H., Goh, H.-N., & Albert, Q. (2024). A review of graduate on time prediction. International Journal on Advanced Science, Engineering and Information Technology, 14(6), 1957–1966.

Mohammad Suhaimi, N., Abdul-Rahman, S., Mutalib, S., Abdul Hamid, N. H., & Md Ab Malik, A. (2019). Review on predicting students' graduation time using machine learning algorithms. International Journal of Modern Education and Computer Science, 11(7), 1–13. https://doi.org/10.5815/ijmecs.2019.07.01

Nanglae, L., Iam-On, N., Boongoen, T., Kaewchay, K., & Mullaney, J. (2021). Determining patterns of student graduation using a bi-level learning framework. Bulletin of Electrical Engineering and Informatics, 10(4), 2201–2211. https://doi.org/10.11591/eei.v10i4.2502

Nguyen, T. K. S., Nguyen, H. Q., & Bui, T. M. (2025). Refining graduation classification accuracy with synergistic deep learning models. Cybernetics and Information Technologies, 25(2), 131–151. https://doi.org/10.2478/cait-2025-0016

Nwankwo, W., Umezuraike, C., Njoku, C. C. (2020). Enhancing Learning Systems using Interactive Intelligent Components. *International Journal of Advanced Trends in Computer Science and Engineering,* 9(3), 3390-3397.

Okoye, K., Nganji, J. T., Escamilla, J., & Hosseini, S. (2024). Computers and Education : Artificial Intelligence

Machine learning model (RG-DMML) and ensemble algorithm for prediction of students ' retention and graduation in education. *Computers and Education: Artificial Intelligence*, *6*(September 2023), 100205. https://doi.org/10.1016/j.caeai.2024.100205

Pawitra, A. S. M., Hung, H.-C., & Jati, H. (2024). A Machine Learning Approach to Predicting On-Time Graduation in Indonesian Higher Education A Machine Learning Approach to Predicting On-Time Graduation in Indonesian Higher Education. *Elinvo (Electronics , Informatics , and Vocational Education)*, *9*(2), 294–308.

Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting University Student Graduation Using Academic Performance and Machine Learning : A Systematic Literature Review. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2024.3361479

Rabelo, A. M., & Zarate, L. E. (2025). A model for predicting dropout of higher education students. *Data Science and Management*, *8*(July 2024), 72–85. https://doi.org/10.1016/j.dsm.2024.07.001

Riadi, I., Umar, R., & Anggara, R. (2024). Comparative analysis of Naive Bayes and K-NN approaches to predict timely graduation using academic history. International Journal of Computing and Digital Systems, 16(1), 1163–1174. https://doi.org/10.12785/ijcds/160185

Sadqui, A., Ertel, M., Sadiki, H., & Amali, S. (2023). Evaluating machine learning models for predicting graduation timelines in Moroccan universities. International Journal of Advanced Computer Science and Applications, 14(7), 304–310. https://doi.org/10.14569/IJACSA.2023.0140734

Sarker, I. H. (2021). Machine Learning : Algorithms , Real - World Applications and Research Directions. *SN Computer Science*, *2*(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

Setiadi, H., Sanjaya, K., Wijayanto, A., Wardhani, D. W., & Cahyono, H. D. (2024). Comparative analysis of classification algorithms using feature selection techniques to predict on-time student graduation. Ingénierie des Systèmes d'Information, 29(4), 1365–1379. https://doi.org/10.18280/isi.290412

Sitio, S. L. M., Darmawati, & Samudra, Y. (2026). Comparative analysis of bagging and boosting models in ensemble learning for graduation prediction. JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 11(3), 979–986. https://doi.org/10.33480/jitk.v11i3.7579

Yağcı, M. (2022). Educational data mining : prediction of students ' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(11), 19. https://doi.org/10.1186/s40561-022-00192-z

Yaqin, A., Rahardi, M., & Abdulloh, F. F. (2022). Accuracy enhancement of prediction method using SMOTE for early prediction student's graduation in XYZ University. International Journal of Advanced Computer Science and Applications, 13(6), 418–424. https://doi.org/10.14569/IJACSA.2022.0130652