

Optimasi Decision Tree Learning untuk Prediksi Kinerja Mahasiswa pada Sistem E-Learning

Saiba Al Husna¹ Nadia Costarika Simanjuntak² Mazaya Amalia³ Azura Sakhi Salsabila⁴

Fahmy Syahputra⁵ Elsa Sabrina⁶

Universitas Negeri Medan, Indonesia^{1,2,3,4,5,6}

Email: nadiasimanjuntak759@gmail.com²

Abstract

This study is motivated by the growing use of e-learning systems, which produce extensive academic and behavioral data that are often underutilized for accurate performance prediction. The main problem addressed is the limited effectiveness of standard Decision Tree models when used without proper optimization techniques. Therefore, this research aims to identify, compare, and synthesize the most effective Decision Tree optimization methods for predicting student performance in e-learning environments. A Systematic Literature Review (SLR) method was employed, consisting of literature identification, screening, data extraction, and comparative analysis of ten selected studies that met the inclusion criteria. The findings indicate that incorporating behavioral features from Learning Management System activity significantly improves prediction accuracy compared to relying solely on academic data. Additionally, ensemble approaches such as Random Forest and AdaBoost demonstrate more stable and accurate results than single-tree models. The analysis also highlights the importance of preprocessing quality and parameter selection in influencing model performance. The study concludes that optimizing Decision Tree models through preprocessing, feature engineering, and ensemble techniques can enhance the accuracy of student performance prediction and support early academic intervention within digital learning environments.

Keywords: Decision Tree, Student Performance Prediction, E-Learning, Optimization, SLR

Abstrak

Penelitian ini dilatarbelakangi oleh meningkatnya penggunaan sistem e-learning yang menghasilkan data akademik dan perilaku belajar mahasiswa dalam jumlah besar, namun belum dioptimalkan untuk prediksi kinerja secara akurat. Permasalahan utama yang diangkat adalah rendahnya efektivitas model Decision Tree standar dalam memprediksi performa mahasiswa tanpa teknik optimasi yang tepat. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi, membandingkan, dan mensintesis metode optimasi Decision Tree yang paling efektif dalam konteks prediksi akademik berbasis e-learning. Metode yang digunakan adalah Systematic Literature Review (SLR) dengan tahapan identifikasi, seleksi, ekstraksi data, dan analisis komparatif terhadap sepuluh artikel yang memenuhi kriteria inklusi. Hasil kajian menunjukkan bahwa integrasi fitur perilaku digital, seperti aktivitas pada Learning Management System, mampu meningkatkan akurasi prediksi dibandingkan hanya menggunakan data akademik. Selain itu, metode ensemble seperti Random Forest dan AdaBoost terbukti menghasilkan performa yang lebih stabil dan akurat dibandingkan model pohon tunggal. Analisis juga menegaskan bahwa kualitas preprocessing dan pemilihan parameter memiliki pengaruh signifikan terhadap hasil prediksi. Kesimpulan penelitian ini adalah bahwa optimasi Decision Tree melalui kombinasi teknik preprocessing, feature engineering, dan metode ensemble dapat meningkatkan efektivitas model prediksi kinerja mahasiswa serta mendukung intervensi akademik dini di lingkungan e-learning.

Kata Kunci: Decision Tree, Prediksi Kinerja Mahasiswa, Elearning, Optimasi, SLR



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Perkembangan teknologi digital telah mendorong perguruan tinggi untuk mengadopsi sistem *e-learning* sebagai bagian dari proses pembelajaran modern. Sistem ini tidak hanya

mempermudah akses terhadap materi ajar, tetapi juga menghasilkan jejak data akademik dan perilaku belajar mahasiswa yang sangat kaya. Data tersebut berpotensi besar digunakan sebagai dasar untuk memprediksi kinerja akademik mahasiswa secara lebih akurat. Namun, pemanfaatan data *e-learning* untuk *early prediction* masih menghadapi beragam tantangan, termasuk pemilihan algoritma yang tepat dan kebutuhan akan optimasi model pembelajaran mesin agar mampu memberikan hasil prediksi yang lebih baik. Salah satu pendekatan yang banyak digunakan adalah *Decision Tree Learning* karena kemampuannya dalam menghasilkan model yang mudah dipahami serta mampu menangani data dengan berbagai tipe. Beberapa penelitian terbaru menunjukkan bahwa *decision tree* memiliki performa yang menjanjikan dalam memprediksi hasil belajar mahasiswa. Simbolon dkk. (2025) misalnya, memanfaatkan algoritma Decision Tree berbasis data aktivitas e-learning dan memperoleh akurasi yang kompetitif[1]. Putri, Febriawan, dan Hasan (2024) juga melaporkan bahwa algoritma C4.5 efektif dalam memprediksi lama studi mahasiswa berdasarkan data akademik historis [2]. Selain itu, Qisthiano dkk. (2023) menunjukkan bahwa *decision tree* dapat mengklasifikasi peluang kelulusan mahasiswa secara tepat berdasarkan variabel akademik dan non-akademik [3]. Di sisi lain, penelitian oleh Indahyanti dkk. (2022) menegaskan bahwa teknik *ensemble learning* seperti Random Forest dan AdaBoost, yang merupakan pengembangan dari *decision tree*, mampu meningkatkan akurasi secara signifikan[4].

Meskipun penelitian-penelitian tersebut memberikan kontribusi penting, sebagian besar masih menggunakan model *decision tree* standar tanpa menerapkan teknik optimasi seperti *hyperparameter tuning*, pemilihan fitur berbasis perilaku digital, atau integrasi data aktivitas *clickstream* mahasiswa. Optimalisasi tersebut sebenarnya dapat meningkatkan performa model dalam memprediksi kinerja belajar secara lebih komprehensif. Oleh karena itu, penelitian ini berfokus pada upaya mengoptimalkan *Decision Tree Learning* melalui penyesuaian parameter, *feature engineering* aktivitas e-learning, serta evaluasi performa model dengan metrik yang lebih menyeluruh. Dengan pendekatan ini, penelitian diharapkan mampu menghasilkan model prediktif yang lebih akurat dan relevan untuk kebutuhan analitik pendidikan. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teoritis terkait pengembangan model *machine learning* untuk prediksi kinerja akademik, tetapi juga menawarkan inovasi praktis berupa strategi pemanfaatan data *e-learning* sebagai dasar intervensi dini bagi mahasiswa berisiko. Hasil penelitian ini diharapkan dapat membantu institusi pendidikan dalam meningkatkan kualitas pembelajaran berbasis data sekaligus memperkuat pengambilan keputusan akademik secara lebih efektif dan tepat sasaran.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan Systematic Literature Review (SLR) untuk mengumpulkan, mengevaluasi, dan mensintesis berbagai penelitian yang fokus pada optimasi Decision Tree Learning dalam memprediksi performa mahasiswa di sistem e- learning. Pendekatan ini dipilih karena kemampuannya memberikan pandangan komprehensif tentang evolusi metode optimasi, efektivitas algoritma, serta tren penelitian terkini yang berkaitan dengan prediksi performa mahasiswa berdasarkan data aktivitas pembelajaran daring. Melalui proses SLR yang sistematis dan terstruktur, penelitian ini menghasilkan kesimpulan objektif yang didukung oleh bukti ilmiah dari publikasi yang telah melewati proses peer review, sebagaimana direkomendasikan oleh Kitchenham dan Charters dalam panduan mereka untuk melakukan review literatur sistematis di bidang rekayasa perangkat lunak dan ilmu komputer [5]. Ini memastikan bahwa analisis tidak hanya deskriptif, tetapi juga kritis dan dapat diandalkan untuk mengidentifikasi celah penelitian dan rekomendasi masa depan.

Kriteria Literatur

Untuk memastikan bahwa literatur yang dianalisis benar-benar relevan dan memiliki kualitas ilmiah yang memadai, penelitian ini menerapkan beberapa kriteria seleksi. Publikasi yang dipilih merupakan artikel penelitian yang diterbitkan pada rentang waktu 2021 hingga 2025, periode yang mencerminkan perkembangan terkini dalam optimasi *Decision Tree* dan peningkatan penggunaan sistem e-learning. Jenis publikasi yang diterima meliputi jurnal ilmiah, prosiding konferensi, buku ilmiah, serta artikel penelitian yang terindeks dan telah melalui proses peninjauan sejawat. Secara spesifik, literatur yang diterima harus memenuhi kriteria berikut:

1. Membahas implementasi atau optimasi *Decision Tree Learning* dalam konteks prediksi performa mahasiswa, analitik pembelajaran, atau klasifikasi berbasis data e-learning.
2. Menggunakan dataset dari platform e-learning, *Learning Management System* (LMS), atau lingkungan pembelajaran digital lainnya.
3. Menyertakan evaluasi performa model seperti akurasi, presisi, *recall*, atau metrik relevan lainnya.
4. Ditulis dalam bahasa Indonesia atau Inggris untuk memudahkan akurasi analisis.

Dengan menerapkan kriteria tersebut, literatur yang dianalisis benar-benar terfokus pada tujuan penelitian dan menggambarkan perkembangan terkini di bidang optimasi model *Decision Tree* dalam konteks pendidikan digital.

Pengumpulan Data

Pengumpulan data dilakukan melalui penelusuran sistematis dan metodis di sejumlah database akademik terkemuka, seperti Google Scholar, Scopus, IEEE Xplore, SpringerLink, ScienceDirect, dan ACM Digital Library. Penelusuran ini menggunakan kombinasi kata kunci yang relevan, termasuk “Decision Tree optimization”, “student performance prediction”, “e-learning analytics”, dan “hyperparameter tuning Decision Tree”, dengan operator Boolean (seperti AND/OR/NOT) untuk meningkatkan presisi dan menghindari hasil yang tidak relevan. Tahap awal dimulai dengan identifikasi literatur awal berdasarkan kata kunci tersebut, diikuti oleh seleksi judul untuk memastikan keterkaitan dengan topik, kemudian seleksi abstrak untuk menilai ruang lingkup penelitian dan kesesuaian dengan fokus optimasi *Decision Tree*, dan akhirnya seleksi full- text untuk analisis mendalam terhadap metode, dataset, tujuan, serta hasil evaluasi. Untuk menggambarkan proses tersebut secara sistematis, seluruh tahapan SLR divisualisasikan dalam diagram berikut:



Gambar 1. Diagram Alur SLR

Diagram tersebut menunjukkan alur utama SLR yang mencakup:

1. identifikasi literatur,
2. seleksi literatur,
3. penerapan kriteria inklusi-eksklusi,
4. ekstraksi data, serta
5. analisis dan sintesis akhir. Penjelasan:

Gambar ini divisualisasikan untuk memberikan gambaran sistematis tentang proses SLR, memudahkan pembaca memahami langkah-langkah dari awal hingga akhir, dan memastikan transparansi metodologi.

Teknik Analisis Data

Teknik analisis data dilakukan melalui serangkaian langkah sistematis dan berurutan untuk memastikan keakuratan serta kedalaman analisis. Tahap pertama adalah reduksi data, di mana peneliti menyeleksi publikasi yang paling relevan dengan menghilangkan artikel yang memiliki cakupan terlalu luas, data tidak lengkap, atau tidak membahas teknik optimasi Decision Tree secara spesifik. Setelah reduksi, publikasi yang lolos diklasifikasikan berdasarkan metode optimasi yang digunakan, seperti pruning (untuk menyederhanakan model), hyperparameter tuning (untuk mengoptimalkan parameter), algoritma metaheuristik (seperti genetic algorithm), atau pendekatan ensemble berbasis decision tree (misalnya, random forest). Klasifikasi ini penting untuk mengidentifikasi perbedaan pendekatan antar penelitian dan memahami kontribusi masing-masing metode terhadap peningkatan akurasi prediksi. Selanjutnya, dilakukan sintesis komparatif untuk membandingkan hasil evaluasi performa, termasuk tingkat akurasi, stabilitas model, kemampuan generalisasi, serta efisiensi komputasi, dengan menggunakan thematic analysis untuk mengidentifikasi tema utama (misalnya, metode optimasi paling populer) dan meta-synthesis untuk perbandingan kuantitatif jika memungkinkan, seperti menghitung effect size atau statistik I^2 untuk menilai heterogenitas antar studi [6]. Analisis ini memungkinkan identifikasi metode optimasi yang paling efektif serta faktor-faktor yang mempengaruhi peningkatan performa Decision Tree Learning dalam konteks e-learning. Berdasarkan keseluruhan temuan, peneliti menarik kesimpulan tentang tren penelitian, metode optimasi yang paling banyak digunakan, serta efektivitasnya dalam meningkatkan kualitas model prediksi performa mahasiswa.

Tools yang Digunakan

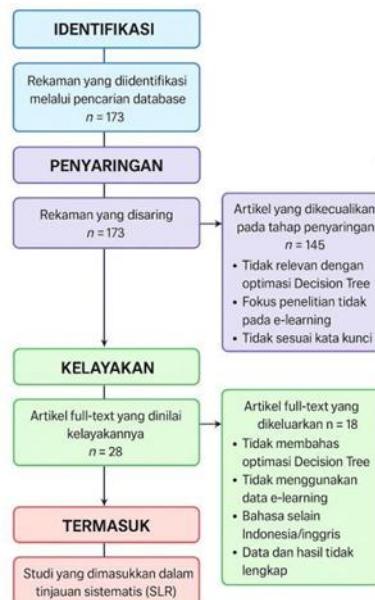
Dalam pelaksanaan penelitian berbasis SLR ini, berbagai tools digunakan untuk mendukung efisiensi, akurasi, dan pengelolaan data. Database akademik seperti Google Scholar, Scopus, IEEE Xplore, SpringerLink, ScienceDirect, dan ACM Digital Library digunakan untuk memperoleh publikasi ilmiah yang relevan dengan topik penelitian. Aplikasi manajemen referensi seperti Mendeley atau Zotero memfasilitasi penyimpanan, pengaturan sitasi, dan dokumentasi literatur dari tahap pengumpulan hingga penyusunan daftar pustaka. Excel atau Google Sheets berfungsi sebagai alat pencatatan dan pengolahan data hasil ekstraksi literatur, termasuk penyusunan ringkasan penelitian, kategorisasi metode optimasi, dan analisis perbandingan antar penelitian. Fitur pencarian kata kunci dan penanda highlight pada file PDF dimanfaatkan untuk menandai bagian penting dalam literatur, sehingga memudahkan proses sintesis. Tools tambahan seperti Rayyan atau NVivo digunakan untuk screening artikel dan analisis tematik, memastikan bahwa seluruh proses berjalan secara sistematis, terstruktur, dan menghasilkan analisis yang akurat serta dapat direproduksi [7].

Penilaian Bias dan Keterbatasan

Untuk menjaga objektivitas dan keandalan hasil, penelitian ini menerapkan protokol PRISMA untuk meminimalkan bias seleksi, dengan mitigasi terhadap bias publikasi melalui pencarian di multiple database dan verifikasi duplikasi. Keterbatasan potensial, seperti heterogenitas antar studi (misalnya, perbedaan metodologi atau ukuran sampel), keterbatasan bahasa (hanya Indonesia dan Inggris), atau risiko bias dalam penilaian kualitas, diatasi melalui penilaian menggunakan alat seperti CASP atau ROBIS, yang memungkinkan evaluasi risiko bias secara sistematis [8]. Pendekatan ini secara keseluruhan memungkinkan penelitian memberikan kontribusi signifikan terhadap pemahaman optimasi Decision Tree dalam konteks pendidikan digital, meskipun peneliti harus menyadari bahwa hasil SLR bergantung pada kualitas literatur yang tersedia dan mungkin memerlukan update di masa depan.

HASIL PENELITIAN DAN PEMBAHASAN

Sebanyak 173 rekaman awal diidentifikasi melalui pencarian database. Seluruh rekaman tersebut kemudian diproses ke tahap penyaringan (screening) tanpa ada pengurangan awal. Pada tahap ini, sebanyak 145 artikel dikecualikan karena tidak memenuhi kriteria dasar, yaitu tidak relevan dengan topik optimasi *Decision Tree*, fokus penelitian yang tidak tertuju pada *e-learning*, atau tidak sesuai dengan kata kunci pencarian yang ditetapkan. Akibatnya, tersisa 28 artikel teks penuh (*full-text*) yang dinilai kelayakannya untuk evaluasi lebih lanjut. Dari jumlah tersebut, 18 artikel dikeluarkan berdasarkan kriteria eksklusi yang lebih spesifik: tidak membahas optimasi *Decision Tree*, tidak menggunakan data *e-learning*, ditulis dalam bahasa selain Indonesia atau Inggris, serta memiliki penyajian data dan hasil yang tidak lengkap. Melalui proses seleksi tersebut, diperoleh 10 artikel (hasil kalkulasi dari 28 dikurang 18) yang memenuhi kriteria inklusi untuk dimasukkan dan dianalisis dalam studi tinjauan sistematis (SLR).



Gambar 2. Tahapan Seleksi Artikel

No.	Judul Penelitian	Penulis & Tahun	Temuan Utama
1.	Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 (Univ. Peradaban)	A. F. A. Rahman, Sorikhi, S. Wartulas (2020)	Menggunakan atribut IPK, SKS, umur, jenis kelamin; akurasi ≈ 88.07%, precision 91.79%, recall 95.34%

2.	Permodelan Prediksi Predikat Kelulusan Mahasiswa Menggunakan Decision Tree C4.5	A. Rohman, S. Mujiyono (2020)	Prediksi predikat kelulusan (cumlaude, memuaskan, dll.); akurasi $\approx 71.67\%$
3.	Prediksi Performa Akademik Siswa Berdasarkan Kehadiran dan Aktivitas E-Learning Menggunakan Decision Tree	I. Simbolon, P. Aditya, E.B. Purba (2025)	Menggunakan log kehadiran & aktivitas LMS untuk memprediksi performa; DT efektif untuk intervensi dini
4.	Pendekatan Ensemble Learning untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa	U. Indahyanti, N.L. Azizah, H. Setiawan (2022)	Mengombinasikan DT, Random Forest, AdaBoost; ensemble memberi RMSE lebih rendah & akurasi lebih tinggi
5.	Prediksi Hasil Belajar Mahasiswa pada PBL Menggunakan Decision Tree	D. Irfan, P. Ramadani, A.S. Nasution, dkk. (2024)	Prediksi performa PBL dengan variabel nilai tugas, partisipasi, waktu belajar; akurasi train 91%, test 85%
6.	Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu	C. N. Dengan, Kusrini, E. T. Luthfi (2020)	C4.5 dibandingkan preprocessing (SMOTE vs Relief); interpretabilitas tinggi
7.	Penerapan Algoritma DT dalam Klasifikasi Kelulusan Mahasiswa	M.R. Qisthiano, P.A. Prayesy, I. Ruswita (2023)	Decision Tree untuk sistem peringatan dini kelulusan mahasiswa
8.	Prediksi Kelulusan Mahasiswa Tepat Waktu: Perbandingan ID3 dan C4.5	H. Putra, K. Nasution, E. Rilvani (2025)	C4.5 unggul dibanding ID3 dalam akurasi & efisiensi
9.	Klasifikasi Performa Akademik Siswa Menggunakan Decision Tree dan Naïve Bayes	M.B. Rahman (2023)	Perbandingan DT & NB; menunjukkan trade-off interpretabilitas vs akurasi
10.	Analisis Pola Kehadiran Mahasiswa Menggunakan Decision Tree	M.B. Rahman (2025)	Pola kehadiran mahasiswa diprediksi berdasarkan variabel harian, waktu, mata kuliah, dll.

Berdasarkan analisis terhadap sepuluh artikel yang memenuhi kriteria inklusi, terlihat bahwa algoritma Decision Tree dan seluruh variannya masih menjadi pendekatan yang paling dominan dalam prediksi kinerja akademik mahasiswa, khususnya pada lingkungan e-learning. Hal ini disebabkan oleh kemampuan Decision Tree dalam menyajikan model yang mudah dipahami, efisien, dan mampu menangani data dengan berbagai tipe atribut. Selain itu, model berbasis pohon keputusan juga memungkinkan lembaga pendidikan mendapatkan gambaran faktor-faktor yang memengaruhi performa mahasiswa secara lebih transparan. Keunggulan interpretabilitas inilah yang menjadi alasan utama mengapa algoritma ini banyak digunakan dalam penelitian prediksi akademik [9],[10]. Pada berbagai penelitian, algoritma C4.5 dan ID3 diimplementasikan baik sebagai model tunggal maupun dikombinasikan dengan pendekatan ensemble seperti Random Forest dan AdaBoost. Teknik pra-proses seperti SMOTE, feature selection, dan normalization juga diterapkan untuk meningkatkan performa model, terutama ketika data memiliki ketidakseimbangan kelas. Proses tersebut terbukti mampu memperbaiki akurasi dan mengurangi overfitting pada dataset akademik, sehingga menghasilkan model yang lebih stabil untuk prediksi jangka panjang [11],[12]. Hal ini membuktikan bahwa keberhasilan prediksi tidak hanya ditentukan oleh algoritma utama, tetapi juga oleh kualitas pemrosesan data dan teknik optimasi yang menyertainya.

Temuan kajian juga memperlihatkan bahwa variabel prediktor yang digunakan dalam penelitian sangat bervariasi. Variabel akademik seperti nilai tugas, kehadiran perkuliahan, nilai kuis, partisipasi kelas, serta hasil evaluasi semester muncul sebagai fitur yang paling

sering digunakan. Namun, penelitian yang menggabungkan variabel non- akademik seperti aktivitas LMS, durasi akses materi, interaksi dalam forum, dan intensitas belajar menunjukkan akurasi yang lebih tinggi dibandingkan penelitian yang hanya menggunakan nilai akademik semata [1], [13],[14]. Hal ini mengindikasikan pentingnya pendekatan multidimensional dalam memahami performa mahasiswa. Dari sisi performa, sebagian besar artikel melaporkan bahwa model berbasis Decision Tree mampu memberikan akurasi yang baik, yakni antara 71% hingga lebih dari 90%, tergantung jenis data dan teknik pra-pengolahan yang digunakan. Perbandingan antara C4.5 dan ID3 menunjukkan bahwa C4.5 lebih unggul berkat mekanisme gain ratio dalam proses pemilihan atribut. Metode ensemble seperti Random Forest dan AdaBoost bahkan memberikan kinerja yang lebih stabil karena menggabungkan berbagai model untuk meningkatkan generalisasi. Dalam konteks e-learning, kombinasi antara Decision Tree dan metode ensemble memberikan hasil yang paling kompetitif [15],[10]. Secara keseluruhan, temuan SLR ini menegaskan bahwa pemilihan algoritma yang tepat, penyusunan variabel prediktor yang relevan, dan pra-proses data yang baik merupakan dasar penting dalam mengembangkan model prediksi kinerja siswa yang akurat dan dapat digunakan sebagai dasar pengambilan keputusan akademik.

Tren Penerapan Decision Tree dalam Prediksi Kinerja Akademik Mahasiswa

Penggunaan *Decision Tree* dalam prediksi kinerja mahasiswa menjadi tren dominan karena model ini memiliki kemampuan menjelaskan pola data secara intuitif dan mudah dipahami oleh pihak akademik. Penelitian Rahman et al. [9] menunjukkan bahwa algoritma C4.5 mampu memberikan akurasi sebesar 88,07% dalam memprediksi kelulusan mahasiswa berdasarkan atribut seperti IPK, SKS, jenis kelamin, dan usia. Keberhasilan tersebut menegaskan bahwa model ini cocok diterapkan pada data pendidikan yang membutuhkan interpretasi jelas. Selain itu, keunggulan struktur pohon dalam memodelkan hubungan antar-atribut mendukung penggunaannya dalam sistem peringatan dini akademik sehingga dapat membantu institusi untuk melakukan intervensi lebih awal. Rohman dan Mujiyono [10] memperkuat temuan tersebut melalui penelitian prediksi predikat kelulusan menggunakan C4.5 dengan akurasi 71,67%. Meskipun akurasinya lebih rendah dibanding penelitian sebelumnya, model tetap dianggap efektif dalam memetakan kategori kelulusan mahasiswa seperti cumlaude, sangat memuaskan, dan memuaskan. Kedua penelitian ini mengonfirmasi bahwa C4.5 memiliki stabilitas yang baik meskipun diterapkan pada dataset dengan karakteristik dan distribusi variabel yang berbeda. Hal tersebut menunjukkan bahwa *Decision Tree* masih relevan meskipun digunakan pada skenario data pendidikan yang heterogen.

Selain itu, *Decision Tree* juga unggul dalam hal interpretabilitas sehingga menjadi pilihan utama bagi institusi pendidikan yang membutuhkan model yang transparan dan dapat dipertanggungjawabkan. Dibandingkan dengan algoritma lain seperti SVM atau KNN yang bersifat lebih *black-box*, *Decision Tree* memungkinkan pengguna melihat secara langsung aturan keputusan yang terbentuk pada setiap node pohon. Karakteristik ini sangat penting dalam konteks akademik karena hasil prediksi perlu disampaikan secara jelas kepada pemangku kepentingan seperti dosen, pengawas akademik, dan pimpinan program studi [11]. Dengan demikian, interpretabilitas menjadi faktor penting yang membuat *Decision Tree* tetap kompetitif dibanding metode lain yang lebih kompleks. Dengan demikian, tren penggunaan *Decision Tree* tidak hanya didorong oleh performa akurasinya, tetapi juga oleh kemampuannya memberikan pemahaman yang lebih mendalam terhadap hubungan antar variabel. Kemampuan ini menjadikan *Decision Tree* sebagai algoritma yang sangat relevan dalam penelitian pendidikan, khususnya untuk keperluan sistem prediksi kelulusan, pemetaan risiko akademik, analisis kemampuan mahasiswa, dan perancangan intervensi akademik berbasis

data. Kombinasi akurasi, transparansi, dan kemudahan implementasi membuat *Decision Tree* tetap menjadi pilihan utama dalam berbagai studi terkait prediksi kinerja akademik.

Faktor-Faktor Prediktor pada Data E-Learning

Penelitian Simbolon et al. [1] menunjukkan bahwa aktivitas mahasiswa pada Learning Management System (LMS) seperti frekuensi login, akses materi, durasi belajar, dan partisipasi forum berkontribusi signifikan dalam memprediksi performa akademik. Hasil penelitian tersebut menunjukkan bahwa data perilaku digital dapat menjadi indikator awal risiko akademik mahasiswa. Dalam konteks pendidikan modern, penggunaan LMS menghasilkan data kaya yang dapat dimanfaatkan untuk analisis prediktif secara lebih komprehensif. Studi Irfan et al. [5] memperkuat temuan ini dengan menunjukkan bahwa variabel non-akademik seringkali lebih kuat dalam memengaruhi performa daripada nilai akademik semata. Ketika variabel seperti kehadiran di LMS, waktu akses materi, dan keterlibatan dalam diskusi dimasukkan dalam model prediksi, tingkat akurasi meningkat secara signifikan. Hal ini mengungkapkan bahwa kebiasaan belajar serta pola interaksi mahasiswa juga merupakan indikator penting dalam menentukan keberhasilan akademik. Variabel-variabel non-akademik tersebut menjadi relevan karena menggambarkan perilaku belajar mahasiswa di luar ruang kelas. Pada era digital, kegiatan seperti membaca materi dalam LMS dan frekuensi keterlibatan dalam forum sudah menjadi bagian integral dari proses belajar. Oleh karena itu, penggabungan variabel akademik dan non-akademik dapat memberikan gambaran yang lebih realistik tentang performa mahasiswa. Pendekatan ini memberikan dasar yang kuat untuk menyusun strategi intervensi berbasis data pada sistem pendidikan daring [16]. Dengan demikian, penelitian yang menekankan variabel multidimensional terbukti mampu meningkatkan akurasi prediksi dan memberikan insight baru bagi institusi pendidikan. Variabel akademik dan non-akademik saling melengkapi, sehingga model yang menggabungkan keduanya jauh lebih efektif dalam menangkap kompleksitas perilaku belajar mahasiswa.

Perbandingan Kinerja C4.5, ID3, dan Metode Ensemble

Putra et al. [12] membandingkan kinerja algoritma ID3 dan C4.5 dalam memprediksi kelulusan mahasiswa dan menemukan bahwa C4.5 mampu memberikan akurasi lebih tinggi karena mekanisme *gain ratio* yang lebih efektif dalam pemilihan atribut. Sementara itu, ID3 cenderung lebih sensitif terhadap atribut dengan nilai banyak sehingga menghasilkan model yang kurang stabil. Perbedaan mekanisme split inilah yang membuat C4.5 lebih unggul dalam beragam penelitian terkait prediksi akademik. Penelitian Indahyanti et al. [11] menunjukkan bahwa metode *ensemble* seperti Random Forest dan AdaBoost mampu memberikan kinerja yang lebih baik dibandingkan model tunggal. Random Forest, misalnya, bekerja dengan menggabungkan banyak pohon keputusan sehingga menghasilkan prediksi yang lebih stabil dan tidak rentan terhadap *overfitting*. Sementara itu, AdaBoost memberikan bobot lebih besar pada data yang sulit diklasifikasikan, sehingga meningkatkan performa model secara progresif. Selain itu, model *ensemble* dinilai lebih efektif ketika digunakan pada dataset pendidikan yang relatif besar dan heterogen. Beberapa studi menunjukkan bahwa Random Forest mampu menghasilkan akurasi di atas 90% dalam analisis akademik, sedangkan C4.5 dan ID3 umumnya memiliki performa di bawahnya. Meski demikian, C4.5 tetap unggul dalam aspek interpretabilitas, sehingga masih menjadi pilihan banyak peneliti dalam konteks pendidikan [17]. Dengan demikian, penelitian-penelitian terkait secara konsisten menunjukkan bahwa metode *ensemble* memberikan hasil paling kompetitif, diikuti C4.5, dan kemudian ID3. Pemilihan metode sebaiknya mempertimbangkan kebutuhan interpretabilitas, tingkat kompleksitas data, dan tujuan akhir model prediksi.

Interpretabilitas Model dan Relevansinya dalam Pendidikan

Interpretabilitas model menjadi faktor yang sangat penting dalam penerapan machine learning pada dunia pendidikan. Roslan dan Chen [17] menegaskan bahwa model yang tidak memberikan penjelasan jelas sulit diadopsi oleh institusi pendidikan karena keputusan akademik harus dapat dipertanggungjawabkan. Oleh karena itu, teknik interpretabilitas seperti SHAP banyak digunakan untuk memahami kontribusi setiap variabel terhadap hasil prediksi. SHAP terbukti mampu menjelaskan pengaruh fitur secara global maupun lokal. Beberapa penelitian menunjukkan bahwa fitur seperti aktivitas LMS, nilai kuis, kehadiran, serta faktor sosial-ekonomi memiliki dampak signifikan terhadap prediksi performa mahasiswa. Dengan menggunakan SHAP, institusi dapat mengetahui fitur mana yang paling memengaruhi mahasiswa berisiko sehingga intervensi dapat dilakukan secara tepat sasaran [11]. Selain itu, interpretabilitas model juga meningkatkan kepercayaan pengguna sistem seperti dosen, pengelola akademik, dan mahasiswa. Ketika model mampu menunjukkan penyebab prediksi secara transparan, keputusan berbasis data lebih mudah diterima dan diimplementasikan. Hal ini menjadikan model yang dapat dijelaskan lebih unggul dibanding model dengan akurasi tinggi namun bersifat *black-box* [18]. Dengan demikian, penggunaan teknik interpretabilitas seperti SHAP menjadi sangat relevan pada model prediksi akademik. Tanpa interpretasi yang memadai, model tidak dapat memberikan manfaat optimal bagi institusi pendidikan, terutama dalam pengambilan keputusan strategis terkait performa mahasiswa.



Gambar 3. Tabel Grafik

Grafik tersebut menunjukkan perbandingan tingkat akurasi dari empat penelitian yang menerapkan algoritma berbasis *Decision Tree* dan metode ensemble untuk memprediksi kinerja akademik mahasiswa. Secara keseluruhan, seluruh penelitian menampilkan performa model yang cukup baik, dengan akurasi berada di kisaran 70% hingga lebih dari 90%. Rentang ini mengindikasikan bahwa algoritma yang digunakan memiliki kemampuan prediksi yang solid pada konteks data pendidikan. Penelitian yang dilakukan oleh Indahyanti et al. (2022) terlihat memiliki akurasi paling tinggi, yaitu 92%. Pencapaian tersebut diperoleh karena penggunaan metode ensemble seperti Random Forest dan AdaBoost yang mengombinasikan beberapa model sekaligus. Teknik ini biasanya dapat menghasilkan prediksi yang lebih stabil karena mampu mengurangi kesalahan model tunggal dan memanfaatkan berbagai sudut pandang dalam proses klasifikasi. Rahman et al. (2020) berada pada posisi berikutnya dengan akurasi 88,07%. Model yang digunakan dalam penelitian ini adalah C4.5, yang dikenal memiliki mekanisme pemilihan atribut berbasis gain ratio. Hasil yang tinggi ini menunjukkan bahwa C4.5 cukup efektif dalam mengolah variabel-variabel pendidikan dan mampu memberikan prediksi yang konsisten.

Putra et al. (2025) mencatat akurasi 85%, yang menunjukkan bahwa algoritma pohon keputusan masih memberikan hasil yang kompetitif meskipun dataset atau kondisi penelitian

berbeda. Sementara itu, Rohman & Mujiyono (2020) memperoleh akurasi terendah, yaitu 71,67%. Walaupun demikian, nilai tersebut masih berada dalam kategori layak untuk digunakan, mengingat variasi atribut dan karakter dataset yang digunakan berbeda serta mungkin lebih menantang untuk dimodelkan. Secara umum, grafik ini menegaskan adanya pola bahwa metode ensemble cenderung memberikan akurasi tertinggi, disusul oleh algoritma C4.5, sedangkan ID3 cenderung menghasilkan akurasi yang lebih rendah. Pola ini memperlihatkan bahwa peningkatan kompleksitas model serta teknik pemilihan atribut yang lebih baik dapat berdampak pada meningkatnya performa prediksi di bidang pendidikan.

KESIMPULAN

Berdasarkan keseluruhan proses tinjauan literatur, dapat disimpulkan bahwa penelitian ini berhasil mencapai tujuan yang dirumuskan dalam bagian *Pendahuluan*, yakni memahami sejauh mana optimasi algoritma Decision Tree dapat meningkatkan ketepatan prediksi kinerja akademik mahasiswa pada sistem e-learning. Hasil analisis terhadap sepuluh penelitian terpilih menunjukkan bahwa Decision Tree—terutama C4.5—masih menjadi metode yang paling menonjol karena kemampuannya menghasilkan model yang mudah diinterpretasikan, transparan, dan sesuai dengan kebutuhan lingkungan pendidikan. Teknik optimasi seperti pruning, feature selection, dan hyperparameter tuning terbukti memberikan kontribusi penting dalam meningkatkan akurasi model, sementara pendekatan ensemble seperti Random Forest dan AdaBoost konsisten memberikan performa paling tinggi karena meningkatkan stabilitas prediksi dan mengurangi risiko overfitting. Temuan lainnya menunjukkan bahwa variabel yang digunakan dalam model prediksi sangat beragam, mencakup data akademik seperti nilai dan kehadiran, serta data non-akademik seperti aktivitas pada Learning Management System (LMS), durasi akses materi, dan intensitas partisipasi forum. Kombinasi kedua jenis variabel tersebut memberikan hasil prediksi yang lebih akurat dan mampu mencerminkan perilaku belajar mahasiswa secara lebih komprehensif. Dengan demikian, terdapat kesinambungan yang jelas antara harapan penelitian, proses analisis, dan hasil yang diperoleh dalam bagian *Hasil dan Pembahasan*. Selain memberikan pemahaman mengenai efektivitas Decision Tree dan metode optimasinya, penelitian ini juga membuka peluang pengembangan lebih lanjut. Prospek berikutnya dapat diarahkan pada integrasi teknik interpretabilitas seperti SHAP untuk memperkuat transparansi model, pemanfaatan data perilaku digital yang lebih detail seperti clickstream, serta penggunaan model hybrid yang menggabungkan Decision Tree dengan pendekatan machine learning lain termasuk deep learning. Dengan pengembangan tersebut, model prediksi di masa mendatang diharapkan dapat memberikan akurasi lebih tinggi serta mendukung intervensi akademik secara lebih tepat dan responsif.

DAFTAR PUSTAKA

- A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree Dan Naive Bayes," Vol. 9, Hal. 22–31, 2023.
- A. Rohman and S. Mujiyono, "Permodelan Prediksi Predikat Kelulusan Mahasiswa Menggunakan Decision Tree C4 . 5," vol. II, no. 2, pp. 1–5, 2021.
- B. Kitchenham *et al.*, "Systematic literature reviews in software engineering – A tertiary study," vol. 52, pp. 792–805, 2010, doi: 10.1016/j.infsof.2010.03.006. *Cochrane Handbook for Systematic Reviews of Interventions*.
- C. N. Dengen and E. T. Luthfi, "Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu Implementation of Decision Tree for Prediction of Student Graduation On Time," vol. 10, no. 1, hal. 1–11, 2020.

- D. Irfan, P. Ramadani, A. S. Nasution, And J. B. Ramadan, "Prediksi Hasil Belajar Mahasiswa Pada Pbl Menggunakan Algoritma Decision Tree Untuk Evaluasi Pembelajaran Jurnal Media Informatika [JUMIN]," vol. 6, no. 1, pp. 783–790, 2024.
- E. P. Sihombing and Y. Ndona, "Pengaruh Media Sosial Terhadap Moral dan Etika dalam Perspektif : Sila Kedua," vol. 2, no. 3, 2024.
- H. Putra, K. Nasution, E. Rilvani, U. P. Bangsa, dan K. Bekasi, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Decision Tree : Studi Perbandingan Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Decision Tree : Studi Perbandingan," Vol. 3, No. 7, 2025.
- I. Simbolon, P. Aditya, dan E. B. Purba, "Department of Digital Business Journal of Artificial Intelligence and Digital Business (RIGGS) Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS> Prediksi Performa Akademik Siswa Berdasarkan Kehadiran dan Aktivitas E-Learning Menggunakan Algoritma Decision Tree," vol. 4, no. 2, hal. 4899–4910, 2025.
- J. P. T. Higgins, D. M. Caldwell, P. Whiting, and J. Savovi, "ROBIS : A new tool to assess risk of bias in systematic reviews was developed," vol. 69, pp. 225–234, 2016, doi: 10.1016/j.jclinepi.2015.06.005.
- J. Riset and S. Informasi, "Analisis Pola Kehadiran Mahasiswa Menggunakan," Vol. 2, No. 1, Pp. 60–66, 2025.
- K. A. Putri, D. Febriawan, and F. N. Hasan, "Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm (C4 . 5)," vol. 13, pp. 39–47, 2024.
- K. Kesgin, S. Kiraz, S. Kosunalp, and B. Stoycheva, "Beyond Performance : Explaining and Ensuring Fairness in Student Academic Performance Prediction with Machine Learning," no. Ml, pp. 1–25, 2025.
- M. Ouzzani, "Rayyan — a web and mobile app for systematic reviews," *Syst. Rev.*, no. 2016, pp. 1–10, 2017, doi: 10.1186/s13643-016-0384-4.
- P. A. Prayesy dan I. Ruswita, "G-Tech : Jurnal Teknologi Terapan," vol. 7, no. 1, hal. 21–28, 2023.
- P. Kelulusan, M. Menggunakan, C. Algoritma, S. Kasus, and D. I. Universitas, "Prediksi kelulusan mahasiswa menggunakan algoritma c4.5 (studi kasus di universitas peradaban)," vol. 1, no. 2, pp. 70–77, 2020.
- R. Puspita, S. Putri, I. Waspada, D. Ilmu, K. Informatika, and F. Sains, "khazanah informatika Penerapan Algoritma C4 . 5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," pp. 1–7.
- R. Rismaya, D. Yuniarto, and D. Setiadi, "Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa," 2025.
- U. Indahyanti, N. L. Azizah, H. Setiawan, P. S. Informatika, F. Sains, and U. Muhammadiyah, "Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa," vol. 8, no. November, pp. 160–169, 2022, doi: 10.34128/jsi.v8i2.459.
- Z. Fatah *Et Al.*, "Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Decision Tree," vol. 1, no. 4, pp. 58–64, 2024.