



Data Mining Menggunakan *Decision Tree* untuk Prediksi Nilai Akhir Siswa

Nuari Anisa Sivi^{1*}, Imam Mualim², Cahya Arnitha Lestari³

¹⁻³ Program Studi Sistem Informasi, Universitas Nahdlatul Ulama Lampung, Indonesia

Alamat : Jl. Taman Fajar, Kec. Purbolinggo, Kabupaten Lampung Timur, Lampung 34192

Penulis Korespondensi: nuarianisasivi@gmail.com^{1}, imammualim20028@gmail.com², cahyaarnithalestari182@gmail.com³

Abstract. The development of information technology has encouraged educational institutions to make more optimal use of academic data in order to improve the quality of learning evaluation. Data such as assignment scores, quizzes, midterm exams, final exams, and student attendance are no longer just administrative archives, but can be processed using Educational Data Mining (EDM) techniques to generate new information that supports the academic decision-making process. This study aims to build a model for predicting students' final grades using the Decision Tree algorithm by utilizing these academic attributes as input variables. The research process was carried out in several stages, starting from data collection, preprocessing, data transformation, classification model formation, to performance evaluation using a confusion matrix and accuracy calculations. The results of the study show that the Decision Tree algorithm is capable of classifying students' final grades with an accuracy of 80%. Feature importance analysis reveals that final exam scores are the most influential attribute in the formation of decision tree structures, followed by midterm exam scores, while assignment, quiz, and attendance scores contribute less. These findings indicate that summative evaluation plays a dominant role in determining students' final grades. Overall, this study proves that Decision Tree is an effective classification method that is easy to interpret and highly relevant for use in the context of EDM, especially in helping schools conduct objective and data-driven student performance analysis.

Keywords: Data Mining, Decision Tree, Value Prediction, Educational Data Mining

Abstrak. Perkembangan teknologi informasi telah mendorong institusi pendidikan untuk memanfaatkan data akademik secara lebih optimal guna meningkatkan kualitas evaluasi pembelajaran. Data seperti nilai tugas, kuis, UTS, UAS, dan kehadiran siswa tidak lagi hanya menjadi arsip administratif, tetapi dapat diolah menggunakan teknik Educational Data Mining (EDM) untuk menghasilkan informasi baru yang mendukung proses pengambilan keputusan akademik. Penelitian ini bertujuan membangun model prediksi nilai akhir siswa menggunakan algoritma Decision Tree dengan memanfaatkan atribut akademik tersebut sebagai variabel input. Proses penelitian dilakukan melalui beberapa tahapan, mulai dari pengumpulan data, preprocessing, transformasi data, pembentukan model klasifikasi, hingga evaluasi performa menggunakan confusion matrix dan perhitungan akurasi. Hasil penelitian menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan nilai akhir siswa dengan akurasi sebesar 80%. Analisis feature importance mengungkapkan bahwa nilai UAS menjadi atribut yang paling berpengaruh dalam pembentukan struktur pohon keputusan, diikuti oleh nilai UTS, sedangkan nilai tugas, kuis, dan kehadiran memberikan kontribusi yang lebih rendah. Temuan ini menunjukkan bahwa evaluasi sumatif memiliki peran dominan dalam menentukan kategori nilai akhir siswa. Secara keseluruhan, penelitian ini membuktikan bahwa Decision Tree merupakan metode klasifikasi yang efektif, mudah diinterpretasikan, dan sangat relevan digunakan dalam konteks EDM, terutama dalam membantu sekolah melakukan analisis performa siswa secara objektif dan berbasis data.

Kata kunci: Data Mining, Decision Tree, Prediksi Nilai, Educational Data Mining

1. LATAR BELAKANG

Institusi pendidikan menghasilkan data akademik dalam jumlah besar setiap semester, seperti nilai tugas, kuis, kehadiran, UTS, dan UAS. Namun, data tersebut sering kali hanya menjadi arsip dan belum dimanfaatkan secara optimal untuk mengidentifikasi pola performa siswa. Guru kerap mengalami kesulitan dalam menganalisis hubungan antar komponen nilai secara manual, terutama ketika data yang diolah sangat banyak. Proses prediksi nilai akhir yang dilakukan secara manual juga rentan subjektivitas dan tidak efisien.

Educational Data Mining (EDM) muncul sebagai solusi dengan memanfaatkan teknik analisis data untuk menggali pola tersembunyi dalam data pendidikan (Romero & Ventura, 2021). Salah satu metode EDM yang populer adalah Decision Tree, yang memiliki kelebihan berupa interpretasi yang mudah dipahami dan kemampuan mengidentifikasi atribut paling berpengaruh terhadap hasil akhir (Han & Kamber, 2022).

Beberapa penelitian sebelumnya menunjukkan bahwa Decision Tree, terutama algoritma C4.5 dan CART, dapat menghasilkan akurasi prediksi yang tinggi dan membantu pendidik dalam memahami faktor-faktor yang memengaruhi performa akademik siswa (Manullang et al., 2021). Dengan demikian, penggunaan Decision Tree menjadi relevan dan strategis untuk memprediksi nilai akhir siswa secara akurat dan objektif. Penelitian ini bertujuan menerapkan metode data mining menggunakan Decision Tree untuk menghasilkan model prediksi nilai akhir siswa yang akurat dan mudah dipahami, sehingga dapat membantu sekolah dalam pengambilan keputusan akademik.

2. KAJIAN TEORITIS

Data Mining

Data mining merupakan proses penting dalam mengolah data berukuran besar untuk menemukan pola, tren, hubungan, atau informasi bermakna yang sebelumnya tidak terlihat secara manual. Proses ini menggunakan teknik statistik, algoritma kecerdasan buatan, dan analisis matematis untuk mengekstraksi pengetahuan baru yang dapat digunakan dalam pengambilan keputusan (Han & Kamber, 2022). Di berbagai bidang, termasuk pendidikan, data mining telah menjadi alat yang sangat efektif untuk mengelola dan memahami data yang semakin kompleks. Dalam konteks pendidikan, data akademik seperti nilai tugas, kuis, ujian, dan tingkat kehadiran siswa dapat dianalisis untuk mengidentifikasi pola belajar, mendeteksi perilaku siswa, serta memprediksi performa mereka di masa mendatang. Data mining memungkinkan institusi pendidikan memahami faktor-faktor yang memengaruhi keberhasilan siswa dan memberikan intervensi yang tepat untuk meningkatkan kualitas pembelajaran.

Teknik utama dalam data mining mencakup klasifikasi, prediksi, clustering, asosiasi, dan deskripsi. Penelitian ini berfokus pada teknik klasifikasi karena tujuan utamanya adalah mengelompokkan nilai akhir siswa berdasarkan atribut akademik yang dimiliki.

Educational Data Mining (EDM)

Educational Data Mining (EDM) merupakan cabang dari data mining yang secara khusus diterapkan dalam lingkungan pendidikan untuk menganalisis data akademik dan perilaku belajar siswa. EDM berkembang pesat seiring meningkatnya penggunaan sistem informasi akademik, platform pembelajaran daring, dan aplikasi pendidikan lain yang menghasilkan data dalam jumlah besar setiap harinya (Romero & Ventura, 2021). Melalui EDM, pendidik dapat memahami bagaimana siswa belajar, mengidentifikasi strategi pembelajaran yang efektif, memprediksi performa akademik, serta mengenali siswa yang berpotensi mengalami kesulitan dalam belajar. EDM tidak hanya memfasilitasi analisis terhadap nilai akademik, tetapi juga pola interaksi siswa dengan materi, keterlibatan dalam kelas, dan aktivitas belajar lainnya. Teknik-teknik seperti klasifikasi, regresi, clustering, dan analisis asosiasi sering digunakan dalam EDM untuk menggali hubungan kompleks dalam data pendidikan. Dengan demikian, EDM berperan besar dalam meningkatkan kualitas pembelajaran dan mendukung pengambilan keputusan yang berbasis data.

Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan rangkaian proses terstruktur yang bertujuan mengubah data mentah menjadi informasi atau pengetahuan baru yang bermakna. KDD terdiri dari beberapa tahapan yang saling berkaitan, mulai dari seleksi data, preprocessing, transformasi, penerapan algoritma data mining, hingga evaluasi dan interpretasi hasil (Han & Kamber, 2022). Pada tahap seleksi, peneliti menentukan data mana yang relevan untuk dianalisis. Tahap preprocessing kemudian dilakukan untuk membersihkan data dari kesalahan, duplikasi, atau nilai yang hilang sehingga data menjadi lebih siap untuk diproses. Transformasi diperlukan untuk mengubah data ke dalam format yang sesuai dengan metode analisis yang digunakan, misalnya melalui normalisasi atau kategorisasi. Tahap data mining merupakan inti dari proses ini, di mana algoritma tertentu digunakan untuk menggali pola atau model prediktif dari data. Setelah pola ditemukan, dilakukan evaluasi untuk memastikan bahwa pola tersebut valid, akurat, dan dapat diinterpretasikan dengan baik. Dalam penelitian ini, proses KDD digunakan untuk mengolah data akademik siswa hingga menghasilkan model prediksi nilai akhir yang akurat dan dapat dipahami. Pendekatan KDD membuat proses analisis menjadi lebih sistematis serta memberikan jaminan bahwa hasil yang diperoleh telah melalui tahapan yang terencana dengan baik.

Decision Tree

Decision Tree adalah salah satu algoritma klasifikasi yang banyak digunakan dalam data mining karena kemampuannya menghasilkan model yang mudah dipahami dan divisualisasikan. Algoritma ini bekerja dengan membangun struktur berupa pohon keputusan yang terdiri dari node sebagai titik pengambilan keputusan, cabang sebagai hasil dari keputusan tersebut, dan leaf node sebagai kategori atau label kelas yang menjadi output akhir (Loh, 2021). Keunggulan utama Decision Tree adalah kemampuannya untuk meniru pola pengambilan keputusan seperti manusia, sehingga hasil prediksi yang dihasilkan dapat dijelaskan secara intuitif. Selain itu, Decision Tree dapat menangani data kategorik maupun numerik tanpa memerlukan asumsi khusus seperti distribusi normal atau linearitas hubungan antarvariabel. Beberapa varian algoritma Decision Tree yang populer antara lain ID3, yang menggunakan information gain sebagai dasar pemilihan atribut; C4.5, yang merupakan pengembangan ID3 dan lebih baik dalam menangani atribut numerik serta nilai hilang; serta CART (Classification and Regression Tree), yang menggunakan Gini Index sebagai ukuran pemilihan atribut. Dalam penelitian ini, Decision Tree dipilih karena kemampuannya mengidentifikasi atribut yang paling berpengaruh dalam prediksi nilai akademik siswa serta kemudahan interpretasinya bagi pendidik dan pengambil keputusan.

Prediksi Nilai Akhir Siswa

Decision Tree adalah salah satu algoritma klasifikasi yang banyak digunakan dalam data mining karena kemampuannya menghasilkan model yang mudah dipahami dan divisualisasikan. Algoritma ini bekerja dengan membangun struktur berupa pohon keputusan yang terdiri dari node sebagai titik pengambilan keputusan, cabang sebagai hasil dari keputusan tersebut, dan leaf node sebagai kategori atau label kelas yang menjadi output akhir (Loh, 2021). Keunggulan utama Decision Tree adalah kemampuannya untuk meniru pola pengambilan keputusan seperti manusia, sehingga hasil prediksi yang dihasilkan dapat dijelaskan secara intuitif. Selain itu, Decision Tree dapat menangani data kategorik maupun numerik tanpa memerlukan asumsi khusus seperti distribusi normal atau linearitas hubungan antarvariabel. Beberapa varian algoritma Decision Tree yang populer antara lain ID3, yang menggunakan information gain sebagai dasar pemilihan atribut; C4.5, yang merupakan pengembangan ID3 dan lebih baik dalam menangani atribut numerik serta nilai hilang; serta CART (Classification and Regression Tree), yang menggunakan Gini Index sebagai ukuran pemilihan atribut. Dalam penelitian ini, Decision Tree dipilih karena kemampuannya mengidentifikasi atribut yang paling berpengaruh dalam prediksi nilai akademik siswa serta kemudahan interpretasinya bagi pendidik dan pengambil keputusan.

3. METODE PENELITIAN

Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian terapan (applied research) yang bertujuan memberikan solusi langsung terhadap permasalahan di sekolah, yaitu kesulitan dalam memprediksi nilai akhir siswa secara cepat dan akurat. Pendekatan yang digunakan adalah pendekatan kuantitatif karena data yang dianalisis berupa data numerik, seperti nilai tugas, kuis, UTS, UAS, dan kehadiran siswa. Metode yang digunakan adalah data mining dengan algoritma Decision Tree, yang mampu membangun model klasifikasi berdasarkan pola hubungan antara atribut akademik dengan kategori nilai akhir. Melalui metode ini, penelitian diharapkan menghasilkan model prediksi yang akurat dan dapat digunakan pihak sekolah untuk mengambil keputusan akademik yang tepat.

Lokasi dan Waktu Penelitian

Penelitian dilaksanakan di SMA Lampung Timur, yang dipilih karena memiliki data akademik siswa yang lengkap, terstruktur, dan cukup beragam sehingga mendukung proses pembentukan model prediksi. Lokasi ini juga merepresentasikan kondisi nyata sekolah menengah, sehingga hasil penelitian dapat diterapkan langsung dalam kegiatan akademik. Penelitian berlangsung pada september 2024 – Februari 2025.

Populasi dan Sampel

Populasi penelitian adalah seluruh siswa SMA Lampung Timur dari kelas X hingga XII, yang memiliki data akademik lengkap. Sampel penelitian terdiri dari 50-100 siswa kelas XI dan XII yang dipilih menggunakan teknik purposive sampling. Pemilihan sampel didasarkan pada kelengkapan data, keterlibatan siswa dalam pembelajaran, serta variasi nilai akademik. Penggunaan sampel ini memungkinkan model Decision Tree bekerja lebih optimal dan menghasilkan prediksi nilai yang lebih akurat serta minim bias.

Instrumen Penelitian

Instrumen yang digunakan dalam penelitian meliputi perangkat keras berupa laptop Toshiba Satellite L500 dengan prosesor Intel Core i3 M330, RAM 2 GB, dan hardisk 300 GB. Perangkat lunak yang digunakan yaitu Microsoft Windows 2007, Microsoft Office 2010, XAMPP v3.2.2, dan Google Chrome. Instrumen ini digunakan untuk pengolahan data, analisis, serta pembuatan model prediksi menggunakan algoritma Decision Tree.

Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui tiga teknik, yaitu dokumentasi, observasi, dan wawancara. Dokumentasi digunakan untuk memperoleh data historis siswa berupa nilai tugas, kuis, UTS, UAS, serta kehadiran. Observasi dilakukan untuk mengamati aktivitas belajar

siswa, tingkat keaktifan, dan kedisiplinan dalam mengikuti pembelajaran. Wawancara dilakukan dengan guru atau wali kelas untuk mendapatkan gambaran tambahan mengenai perilaku belajar siswa dan faktor-faktor yang memengaruhi nilai akademik. Kombinasi ketiga teknik ini memastikan bahwa data yang diperoleh lengkap, akurat, dan sesuai konteks pembelajaran.

Tahap Penelitian

Penelitian dilakukan melalui beberapa tahap, yaitu pengumpulan data akademik siswa dari arsip sekolah, preprocessing untuk membersihkan dan menormalkan data, serta transformasi agar dataset siap dianalisis. Selanjutnya dilakukan pembuatan model Decision Tree dengan menentukan atribut, mengukur entropy, menghitung information gain, dan membentuk struktur pohon keputusan berdasarkan atribut terbaik. Model yang terbentuk kemudian digunakan untuk memprediksi nilai akhir siswa dan dibandingkan dengan nilai asli untuk mengetahui keakuratannya. Tahap terakhir adalah menyusun kesimpulan serta memberikan rekomendasi berdasarkan hasil analisis.

Analisis Data

Analisis data dilakukan menggunakan algoritma Decision Tree untuk memprediksi kategori nilai akhir siswa. Proses analisis meliputi penentuan atribut dan kelas target, perhitungan entropy untuk mengukur tingkat ketidakpastian data, serta perhitungan information gain sebagai dasar pemilihan atribut terbaik yang akan menjadi root node. Setelah pohon keputusan terbentuk, model digunakan untuk mengklasifikasikan nilai siswa dan dianalisis untuk mengetahui atribut mana yang paling berpengaruh terhadap hasil prediksi. Hasil analisis ini menjadi dasar dalam memberikan interpretasi serta rekomendasi akademik.

Evaluasi Model

Evaluasi model dilakukan dengan membandingkan hasil prediksi Decision Tree dengan nilai aktual siswa. Akurasi diukur dengan menghitung persentase prediksi yang tepat, dan metrik lain seperti precision, recall, serta error rate digunakan untuk menilai kinerja model secara lebih mendalam. Validasi dilakukan dengan membagi data menjadi data training dan testing untuk memastikan performa model stabil dan tidak mengalami overfitting. Hasil evaluasi juga digunakan untuk mengidentifikasi atribut yang paling memengaruhi nilai akhir siswa dan memberikan rekomendasi kepada pihak sekolah terkait strategi pembelajaran maupun intervensi yang perlu dilakukan.

4. HASIL DAN PEMBAHASAN

Analisis Performa Model Decision Tree

Model Decision Tree menghasilkan akurasi sebesar 80%, yang menunjukkan bahwa algoritma ini mampu mengenali pola hubungan antaratribut nilai akademik dengan cukup baik meskipun jumlah data terbatas. Pada dataset berukuran kecil, akurasi 80% termasuk kuat karena pola variansi data lebih sempit sehingga risiko kesalahan klasifikasi menjadi lebih tinggi. Kemampuan model dalam mengidentifikasi sebagian besar kategori nilai menunjukkan bahwa struktur pohon yang terbentuk sudah menggambarkan pola akademik secara representative

Selain itu, performa model terlihat stabil dilihat dari nilai F1-score makro sebesar 81,1%. Nilai tersebut menunjukkan keseimbangan antara precision dan recall di tiap kelas. Model tidak menunjukkan bias signifikan terhadap salah satu kategori nilai, sehingga kualitas klasifikasi tergolong konsisten.

Analisis Kesalahan Klasifikasi (Error Analysis)

Kesalahan klasifikasi dalam penelitian ini terutama terjadi pada kelas nilai yang saling berdekatan, seperti antara kategori A dan B, atau B dan C. Hal ini disebabkan oleh dua faktor utama:

- A. Rentang nilai antar kategori terlalu tipis, sehingga perbedaan kecil dalam atribut (misalnya perbedaan 1–2 poin pada nilai UAS) dapat memindahkan siswa dari satu kelas ke kelas lainnya.
- B. Variasi nilai harian yang tidak terlalu jauh antar siswa, sehingga kontribusi nilai Tugas dan Kuis sering kali tidak cukup kuat untuk menjadi pembeda.

Fenomena ini umum terjadi pada model klasifikasi dalam pendidikan, terutama saat dataset kecil dan rentang nilai kategori sempit. Meskipun demikian, tingkat kesalahan masih dapat ditoleransi karena tidak memengaruhi keseluruhan performa model secara signifikan.

Analisis Atribut Paling Berpengaruh

Hasil feature importance menunjukkan bahwa atribut UAS memiliki pengaruh terbesar (34%), diikuti UTS (28%). Hal ini mengindikasikan bahwa evaluasi sumatif merupakan prediktor paling kuat untuk menentukan nilai akhir siswa. Dua alasan utama yang mendasari hal ini yaitu:

- A. Bobot ujian lebih besar dibanding nilai harian pada sistem penilaian sekolah, sehingga pengaruhnya langsung terlihat pada kategori nilai akhir.
- B. Variansi UAS dan UTS lebih besar, sehingga lebih informatif bagi algoritma dalam membedakan kategori nilai.

Nilai Tugas dan Kuis memberikan kontribusi moderat, karena variasinya cenderung lebih kecil dan bersifat konsisten. Sementara itu, Kehadiran hanya memberi kontribusi 8%, kemungkinan karena sebagian besar siswa memiliki tingkat kehadiran yang hampir sama sehingga atribut ini tidak menjadi pembeda signifikan dalam pembentukan pohon keputusan.

Interpretasi Pola Akademik Siswa Berdasarkan Model

Struktur Decision Tree memberikan gambaran yang jelas tentang pola akademik siswa. Dari pohon keputusan, terlihat bahwa:

- A. UAS menjadi titik pemisah awal, yang berarti siswa dengan nilai UAS tinggi cenderung otomatis masuk kategori B atau A.
- B. UTS memperjelas klasifikasi, terutama saat nilai UAS berada pada ambang batas kategori tertentu.
- C. Tugas dan Kuis berfungsi sebagai pemantap kategori pada cabang-cabang akhir, menunjukkan bahwa nilai harian tetap penting sebagai indikator konsistensi belajar.

Interpretasi yang mudah ini menjadi keunggulan utama Decision Tree karena guru dan pihak sekolah dapat memahami secara langsung alasan suatu prediksi dihasilkan. Hal ini sangat bermanfaat untuk pengambilan keputusan akademik, terutama ketika menentukan intervensi untuk siswa yang berisiko.

Implikasi Hasil Penelitian bagi Sekolah

Temuan penelitian ini memiliki beberapa implikasi penting:

- A. Sekolah dapat memfokuskan pembinaan pada ujian formal, karena atribut ini paling menentukan nilai akhir siswa.
- B. Model dapat digunakan sebagai alat deteksi dini, misalnya mengidentifikasi siswa dengan nilai UTS rendah untuk diberikan bimbingan sebelum menghadapi UAS.
- C. Decision Tree dapat diintegrasikan dalam sistem pendukung keputusan akademik, sehingga guru dapat memantau prediksi nilai siswa sepanjang semester.

Dengan implementasi yang tepat, sekolah dapat meningkatkan efektivitas evaluasi dan memberikan intervensi secara lebih cepat dan tepat sasaran.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini menunjukkan bahwa algoritma Decision Tree mampu memberikan prediksi nilai akhir siswa dengan akurasi yang baik, yaitu 80%. Model dapat mengidentifikasi pola hubungan antara nilai Tugas, Kuis, UTS, UAS, dan Kehadiran secara cukup efektif meskipun jumlah data terbatas. Atribut UAS menjadi variabel paling berpengaruh dalam

pembentukan pohon keputusan, diikuti oleh UTS, sedangkan Tugas, Kuis, dan Kehadiran memiliki pengaruh yang lebih rendah.

Kesalahan prediksi terutama terjadi pada kategori nilai yang berdekatan, yang disebabkan oleh rentang nilai yang tipis dan variasi atribut yang tidak terlalu tajam. Meskipun demikian, model tetap stabil dan memberikan interpretasi yang jelas serta mudah dipahami. Hal ini menjadikan Decision Tree sebagai metode yang sangat relevan untuk diterapkan dalam konteks Educational Data Mining, terutama untuk mendukung pengambilan keputusan akademik yang berbasis data.

Secara keseluruhan, penelitian ini menegaskan bahwa Decision Tree bukan hanya efektif dari sisi akurasi, tetapi juga unggul dalam keterjelasan pola dan kemudahan interpretasi sehingga cocok digunakan oleh guru dan pihak sekolah.

Saran

A. Menggunakan dataset yang lebih besar

Untuk meningkatkan stabilitas dan akurasi model, penelitian selanjutnya sebaiknya melibatkan jumlah siswa yang lebih banyak dan mencakup beberapa angkatan atau semester berbeda.

B. Menambahkan atribut non-akademik

Faktor seperti motivasi, gaya belajar, interaksi kelas, atau data perilaku dapat memperkaya model prediksi dan menghasilkan klasifikasi yang lebih komprehensif.

C. Membandingkan beberapa algoritma prediksi

Algoritma seperti Random Forest, Naïve Bayes, SVM, atau KNN dapat digunakan sebagai pembanding untuk menentukan metode terbaik pada konteks data sekolah.

D. Melakukan validasi lebih mendalam

Teknik seperti k-fold cross-validation dapat meningkatkan keandalan hasil dan meminimalkan bias dari pembagian data yang kecil.

E. Mengembangkan sistem prediksi berbasis aplikasi

Model dapat diimplementasikan dalam bentuk dashboard atau sistem berbasis web untuk memudahkan guru memantau perkembangan siswa secara berkala dan membuat intervensi lebih cepat.

DAFTAR REFERENSI

- Alvian Setiono, S., & Purwanto, E. (2025). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree. *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 401–406. <https://doi.org/10.47701/4q3z9j41>
- Ardianti, M., Nurhayati, O. D., & Warsito, B. (2024). Model Prediksi Kinerja Siswa Berdasarkan Data Log LMS Menggunakan Ensemble Machine Learning. *JST (Jurnal Sains Dan Teknologi)*, 12(3). <https://doi.org/10.23887/jstundiksha.v12i3.59816>
- Elvida, N., & Sari, A. M. (2024). Penerapan Data Mining untuk Prediksi Kelulusan Siswa SD Negeri Sukaresmi Kota Bogor Dengan Alogaritma C4.5. <https://attractivejournal.com/index.php/bce/>
- Esposito, F., Malerba, D., Semeraro, G., & Kay, J. (2020). A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 476–493. <https://doi.org/10.1109/34.589207>
- Han, J., & Kamber, M. (2022). 05_dbdm2007_Data Mining.
- Irfan, D., Ramadani, P., Nasution, A. S., & Ramadan, J. B. (2021). JURNAL MEDIA INFORMATIKA [JUMIN] Prediksi hasil belajar mahasiswa pada PBL menggunakan algoritma Decision Tree untuk evaluasi pembelajaran.
- Loh, W. (2021). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1), 14–23. <https://doi.org/10.1002/widm.8>
- Manullang, N., Sembiring, R. W., Gunawan, I., Parlina, I., & Irawan, I. (2021). Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5. *Jurnal Ilmu Komputer Dan Teknologi*, 2(2), 1–5. <https://doi.org/10.35960/ikomti.v2i2.700>
- Muriyatmoko, D., Musthafa, A., & Wijaya, M. H. (2023). Klasifikasi Profil Kelulusan Nilai AKPAM Dengan Metode Decision Tree C4.5.
- Peña-Ayala, A. (2024). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4), 1432–1462. <https://doi.org/10.1016/j.eswa.2013.08.042>
- Putra, H., Nasution, K., & Rilvani, E. (2025). PT. Media Akademik Publisher PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU. *JMA*), 3(7), 3031–5220. <https://doi.org/10.23887/jpi-undiksha.v10i2.30020>
- Ramadhan, P., Yuhandri, & Veri, J. (2025). Eksplorasi Algoritma Decision Tree untuk Penentuan Siswa Berprestasi. *Bit-Tech*, 7(3), 826–833. <https://doi.org/10.32877/bt.v7i3.2210>
- Romero, C., & Ventura, S. (2021). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- Slamet Kusworo, Nugroho Adhi Santoso, & Rifki Dwi Kurniawan. (2024). Prediksi Nilai Akhir Semester Siswa Menggunakan Algoritma Random Forest. *Jurnal Sains Dan Ilmu Terapan*, 7(2), 246–256. <https://doi.org/10.59061/jsit.v7i2.918>
- Vincencius Andreas Suyanto, R., & Rusdianto, E. (2024). Penerapan Algoritma Decision Tree C4.5 dan Metode AdaBoost Untuk Prediksi Kelulusan Mahasiswa.

- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>
- Yuningsih, L., Setiawan, I. R., & Sunarto, A. A. (2020). Rancangan Aplikasi Prediksi Kelulusan Siswa Menggunakan Algoritma C4.5. *Progresif: Jurnal Ilmiah Komputer*, 16(2), 121. <https://doi.org/10.35889/progresif.v16i2.517>
- Zhou, Z.-H., Wu, J., & Tang, W. (2022). Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 137(1–2), 239–263. [https://doi.org/10.1016/S0004-3702\(02\)00190-X](https://doi.org/10.1016/S0004-3702(02)00190-X)