

Penerapan Data Mining Dalam Prediksi Kinerja Akademik Mahasiswa Menggunakan Algoritma Machine Learning

Vinsensius Yoga Danar Wijaya *¹ Goenawan Brotosaputro ²

¹ Fakultas Teknologi Informasi, Universitas Budi Luhur, Jl. Ciledug Raya - Petukangan Utara, Jakarta Selatan 12260, Indonesia, 2011600729@student.budiluhur.ac.id

² Fakultas Teknologi Informasi, Universitas Budi Luhur, Jl. Ciledug Raya - Petukangan Utara, Jakarta Selatan 12260, Indonesia, goenawan.brotosaputro@budiluhur.ac.id

*Corresponding Author: danaryoga@gmail.com (081382280311)

Abstrak

Prediksi kinerja akademik mahasiswa menjadi topik penting dalam pendidikan tinggi karena dapat membantu institusi mengidentifikasi mahasiswa berisiko dropout. Penelitian ini melakukan Systematic Literature Review (SLR) dengan panduan PRISMA terhadap 527 artikel yang diperoleh dari basis data Scopus, IEEE Xplore, ScienceDirect, SpringerLink, dan Google Scholar (periode 2016–2025). Setelah proses seleksi, 24 artikel memenuhi kriteria inklusi. Hasil analisis menunjukkan bahwa algoritma yang paling banyak digunakan meliputi Random Forest, Decision Tree, Support Vector Machine, Artificial Neural Network, dan Graph Convolutional Networks, dengan tingkat akurasi bervariasi antara 70% hingga 98%. Misalnya, Random Forest mencapai akurasi di atas 90%, sedangkan Decision Tree menunjukkan performa terbaik pada dataset terbatas dengan akurasi 98%. Faktor penentu performa model meliputi kualitas dataset, teknik feature selection, serta metode evaluasi. Penelitian ini menegaskan bahwa meskipun akurasi model terus meningkat, interpretabilitas hasil dan kemampuan generalisasi masih menjadi tantangan. Temuan ini memberikan dasar bagi pengembangan metode prediksi yang lebih efektif dan adaptif di masa depan.

Keywords: Data mining, prediksi kinerja akademik, mahasiswa, machine learning.

Abstract

Predicting student academic performance is an important topic in higher education because it can help institutions identify students at risk of dropping out. This study conducted a Systematic Literature Review (SLR) using the PRISMA guidelines on 527 articles obtained from the Scopus, IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar databases (2016–2025 period). After the selection process, 24 articles met the inclusion criteria. The results of the analysis showed that the most widely used algorithms included Random Forest, Decision Tree, Support Vector Machine, Artificial Neural Network, and Graph Convolutional Networks, with accuracy rates varying between 70% and 98%. For example, Random Forest achieved an accuracy of over 90%, while Decision Tree showed the best performance on a limited dataset with an accuracy of 98%. Factors determining model performance included dataset quality, feature selection techniques, and evaluation methods. This study confirms that although model accuracy continues to improve, the interpretability of results and generalization capabilities remain challenges. These findings provide a basis for the development of more effective and adaptive prediction methods in the future.

Keywords: Data mining, academic performance prediction, students, machine learning.

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan besar dalam berbagai aspek kehidupan, termasuk dalam bidang pendidikan tinggi (Zhang et al., 2021). Perguruan tinggi kini mengelola berbagai bentuk data yang dihasilkan dari aktivitas akademik mahasiswa, seperti nilai ujian, kehadiran, aktivitas di Learning Management System (LMS), dan catatan akademik lainnya. Data tersebut, jika dikelola dan dianalisis dengan tepat, dapat menjadi sumber informasi yang berharga untuk memahami pola belajar dan kinerja mahasiswa. Namun, pengelolaan data dalam jumlah besar sering kali menjadi tantangan tersendiri bagi institusi pendidikan yang belum memiliki sistem analisis terintegrasi (Salal et al., 2019). Oleh karena itu, diperlukan pendekatan yang mampu mengekstrak informasi penting dari

data tersebut secara efisien dan akurat.

Data mining, yang merupakan bagian dari disiplin ilmu Knowledge Discovery in Databases (KDD), hadir sebagai solusi untuk menganalisis data dalam skala besar guna menemukan pola dan hubungan tersembunyi. Dalam konteks pendidikan, data mining dapat digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi kinerja akademik mahasiswa (Saa, 2016). Proses ini melibatkan penggunaan berbagai teknik analisis, seperti klasifikasi, regresi, clustering, dan association rule mining (Namoun & Alshantqi, 2020). Implementasi metode ini tidak hanya membantu pihak universitas memahami perilaku belajar mahasiswa, tetapi juga memungkinkan pengambilan keputusan yang berbasis data (data-driven decision making). Dengan demikian, data mining menjadi alat yang strategis dalam mendukung peningkatan mutu pendidikan.

Prediksi kinerja akademik mahasiswa merupakan salah satu penerapan data mining yang sangat relevan dan bermanfaat bagi perguruan tinggi. Dengan memanfaatkan model prediksi, pihak akademik dapat mengidentifikasi mahasiswa yang berpotensi mengalami kesulitan belajar sebelum masalah tersebut semakin berkembang (Mueen et al., 2016). Misalnya, dengan mengetahui kemungkinan penurunan performa akademik sejak dini, dosen pembimbing dapat memberikan bimbingan atau intervensi khusus. Hal ini tidak hanya berkontribusi pada peningkatan prestasi mahasiswa, tetapi juga dapat membantu menekan angka putus studi (López-Zambrano et al., 2021). Oleh sebab itu, kemampuan memprediksi kinerja akademik menjadi salah satu indikator penting dalam sistem pendidikan yang adaptif.

Dalam praktiknya, penerapan data mining untuk prediksi kinerja akademik sering melibatkan penggunaan algoritma machine learning. Algoritma-algoritma ini, seperti Decision Tree, Random Forest, Support Vector Machine (SVM), dan Artificial Neural Network (ANN), mampu belajar dari data historis untuk membuat prediksi yang lebih akurat (Chaka, 2021). Masing-masing algoritma memiliki kelebihan dan kekurangan, sehingga pemilihan metode yang tepat perlu mempertimbangkan karakteristik data yang digunakan. Selain itu, parameter model yang dioptimalkan dapat memengaruhi tingkat akurasi prediksi. Oleh karena itu, eksperimen komparatif sering dilakukan untuk menentukan algoritma yang paling sesuai untuk konteks tertentu.

Penerapan data mining dalam bidang pendidikan juga selaras dengan tren global menuju smart campus dan learning analytics. Konsep ini menekankan pentingnya penggunaan teknologi untuk mengelola, memantau, dan meningkatkan pengalaman belajar mahasiswa (Kamal & Ahuja, 2017). Melalui analisis data yang komprehensif, universitas dapat merancang strategi pembelajaran yang lebih personalisasi dan efektif. Dengan demikian, prediksi kinerja akademik tidak hanya bermanfaat bagi mahasiswa secara individu, tetapi juga mendukung peningkatan kualitas proses pembelajaran secara keseluruhan (Feng et al., 2022). Hal ini menunjukkan bahwa integrasi data mining ke dalam sistem akademik memiliki nilai strategis jangka panjang.

Selain itu, perkembangan ketersediaan data yang semakin beragam membuka peluang untuk menggabungkan berbagai sumber informasi dalam proses prediksi. Misalnya, selain nilai akademik, data seperti interaksi mahasiswa di platform pembelajaran daring, aktivitas ekstrakurikuler, dan faktor demografis juga dapat dianalisis untuk memberikan gambaran yang lebih holistik (Amrieh et al., 2016). Penggabungan data ini memungkinkan terciptanya model prediksi yang lebih robust dan mampu menangkap kompleksitas faktor-faktor yang memengaruhi prestasi belajar (Alamgir et al., 2024). Hal ini semakin memperkuat argumen bahwa data mining dapat menjadi fondasi penting dalam sistem pendidikan berbasis data.

Dengan semakin meningkatnya kompetisi di dunia pendidikan, perguruan tinggi dituntut untuk terus meningkatkan kualitas layanan akademiknya. Salah satu cara untuk mencapainya adalah dengan memanfaatkan teknologi analisis data yang canggih seperti data mining. Kemampuan untuk memprediksi kinerja akademik mahasiswa secara akurat dapat membantu institusi dalam merencanakan kebijakan yang tepat, menyediakan dukungan akademik yang lebih terarah, dan pada akhirnya menghasilkan lulusan yang lebih kompeten (Ünal, 2020). Oleh karena itu, penelitian mengenai penerapan data mining untuk prediksi kinerja akademik mahasiswa menjadi topik yang relevan dan strategis untuk dikaji lebih lanjut.

Yağci (2022) menerapkan berbagai algoritma machine learning seperti Random Forest, Support Vector Machine, dan Artificial Neural Network untuk memprediksi kinerja akademik mahasiswa berdasarkan data nilai dan kehadiran. Hasilnya menunjukkan bahwa Random Forest memberikan akurasi tertinggi, mencapai lebih dari 90%, sehingga direkomendasikan sebagai metode yang andal untuk prediksi berbasis data pendidikan. Suaza-Medina dkk. (2024) mengembangkan model prediksi kinerja akademik pada wilayah yang tertinggal secara pendidikan menggunakan algoritma Gradient Boosting yang dikombinasikan dengan Shapley Additive Explanations (SHAP). Penelitian ini membuktikan bahwa metode tersebut tidak hanya akurat, tetapi juga memberikan interpretasi yang jelas mengenai kontribusi setiap variabel terhadap hasil prediksi.

Khairy dkk. (2024) menggunakan algoritma Decision Tree, Naive Bayes, dan K-Nearest Neighbor untuk memprediksi hasil ujian mahasiswa berdasarkan data akademik dan perilaku belajar. Hasil penelitian menunjukkan bahwa Decision Tree memiliki tingkat akurasi terbaik, sehingga cocok digunakan dalam sistem peringatan dini (early warning system). Hu dan Rangwala (2020) mengusulkan pendekatan attention-based graph convolutional networks untuk memodelkan hubungan kompleks antar mahasiswa dalam konteks pembelajaran. Model ini terbukti mampu meningkatkan akurasi prediksi kinerja akademik dibandingkan pendekatan berbasis deep learning konvensional.

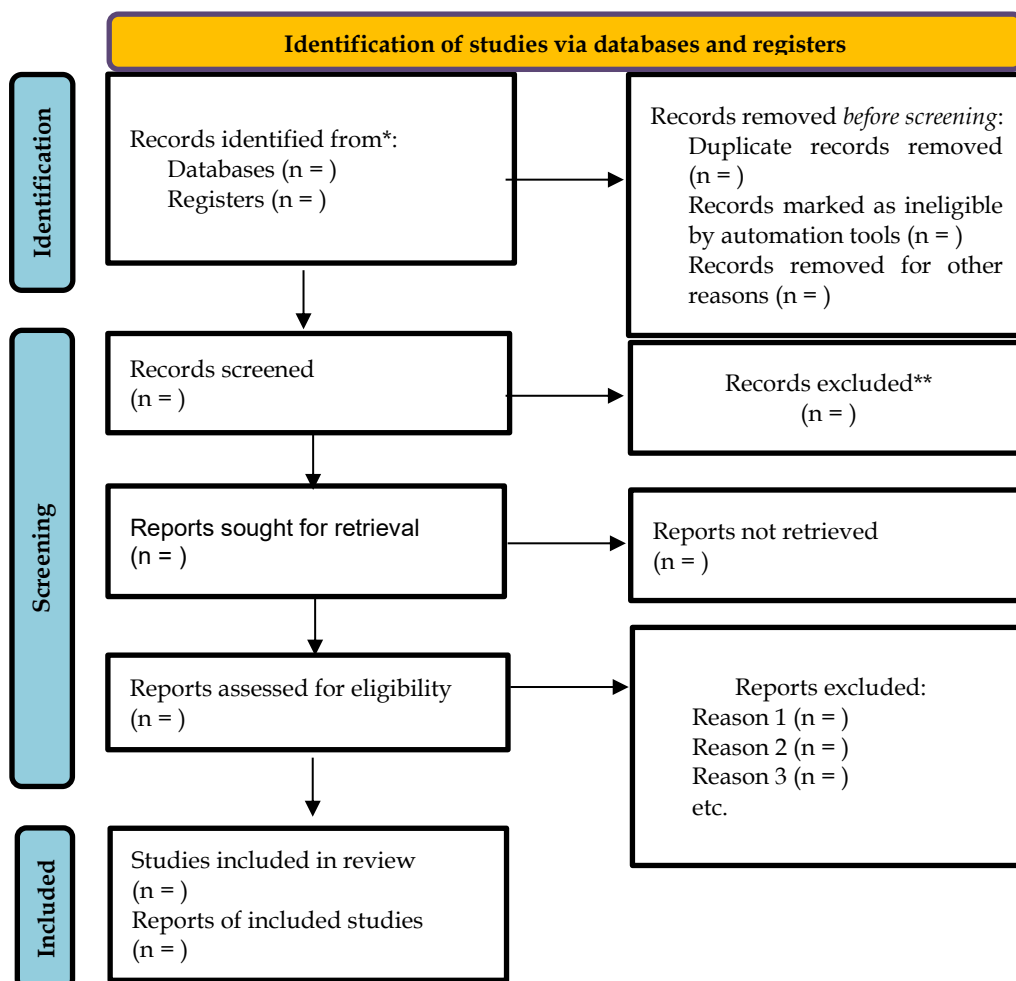
Meskipun penelitian terkait prediksi kinerja akademik mahasiswa menggunakan data mining telah banyak dilakukan, sebagian besar masih berfokus pada penggunaan algoritma tertentu atau studi kasus pada institusi tunggal, sehingga belum memberikan gambaran komprehensif mengenai tren, keunggulan, dan kelemahan berbagai pendekatan yang ada. Selain itu, kajian komparatif yang mengintegrasikan berbagai hasil penelitian terdahulu untuk mengidentifikasi pola umum, faktor penentu keberhasilan, dan peluang pengembangan metode masih terbatas. Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk melakukan Systematic Literature Review (SLR) guna merangkum dan menganalisis secara sistematis temuan-temuan dari penelitian sebelumnya terkait penerapan data mining dalam prediksi kinerja akademik mahasiswa. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam bentuk peta

pengetahuan yang dapat membantu peneliti lain, praktisi pendidikan, dan pengambil kebijakan dalam memilih metode, algoritma, dan variabel yang paling relevan, serta memberikan rekomendasi arah penelitian lanjutan yang potensial.

METODE PENELITIAN

Penelitian ini menggunakan metode Systematic Literature Review (SLR) dengan panduan Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Pendekatan ini dipilih karena mampu mengumpulkan, mengevaluasi, dan mensintesis hasil penelitian secara sistematis sehingga dapat memberikan gambaran menyeluruh mengenai penerapan data mining untuk prediksi kinerja akademik mahasiswa. PRISMA digunakan sebagai kerangka kerja untuk memastikan transparansi, akurasi, dan replikasi dalam proses seleksi dan pelaporan literatur. Proses ini melibatkan empat tahap utama, yaitu identifikasi, penyaringan (screening), penentuan kelayakan (eligibility), dan inklusi. Tahap identifikasi dilakukan dengan pencarian artikel pada basis data ilmiah bereputasi seperti Scopus, IEEE Xplore, ScienceDirect, SpringerLink, dan Google Scholar. Kata kunci yang digunakan antara lain “data mining”, “academic performance prediction”, “student performance”, dan “machine learning in education”, baik secara tunggal maupun dikombinasikan menggunakan operator Boolean (AND, OR). Kriteria inklusi meliputi artikel yang dipublikasikan antara tahun 2016 hingga 2025, berbahasa Inggris atau Indonesia, serta memiliki fokus pada penerapan teknik data mining atau machine learning untuk memprediksi kinerja akademik mahasiswa. Adapun kriteria eksklusi meliputi publikasi berupa tinjauan literatur, artikel tanpa akses penuh, atau penelitian yang tidak relevan dengan fokus kajian. Tahap penyaringan dilakukan dengan menghapus artikel duplikat dan meninjau judul serta abstrak untuk memastikan relevansi dengan topik penelitian. Artikel yang lolos pada tahap ini kemudian diperiksa lebih lanjut pada tahap kelayakan dengan membaca isi penuh untuk menilai kesesuaian dengan kriteria inklusi. Proses ini melibatkan pencatatan informasi penting dari setiap artikel, seperti penulis, tahun publikasi, metode yang digunakan, jenis algoritma, variabel yang dianalisis, dan hasil penelitian. Tahap inklusi dilakukan dengan menentukan artikel yang akan dianalisis secara mendalam untuk sintesis data.

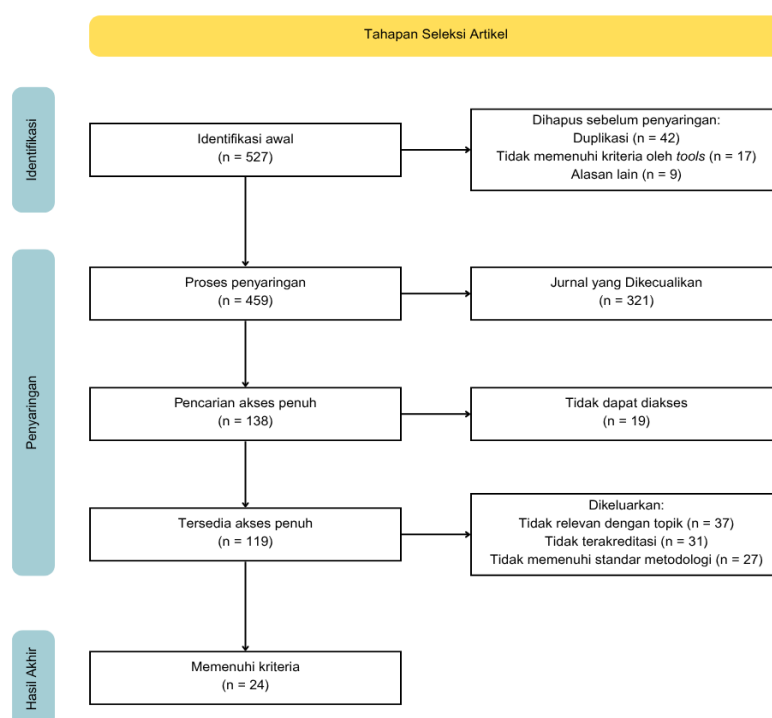
Data dari artikel terpilih dianalisis menggunakan pendekatan sintesis deskriptif, di mana hasil-hasil penelitian dirangkum untuk mengidentifikasi tren, kesenjangan (research gap), serta peluang penelitian di masa depan. Diagram alur PRISMA digunakan untuk menggambarkan proses seleksi artikel mulai dari tahap identifikasi hingga inklusi. Pendekatan ini memungkinkan peneliti menyajikan ringkasan yang terstruktur, akurat, dan dapat dipertanggungjawabkan, sehingga hasil penelitian dapat menjadi acuan yang kuat bagi pengembangan studi lanjutan dalam bidang prediksi kinerja akademik menggunakan data mining.



Gambar 1 Diagram Alur Prisma SLR

HASIL DAN PEMBAHASAN

Sebanyak 527 rekaman awal diidentifikasi dari berbagai basis data ilmiah bereputasi, termasuk Scopus, IEEE Xplore, ScienceDirect, SpringerLink, dan Google Scholar. Sebelum tahap penyaringan, 68 rekaman dihapus, yang terdiri dari 42 duplikat, 17 secara otomatis dikecualikan karena tidak memenuhi kriteria kelayakan awal, dan 9 dihapus akibat masalah teknis atau kesalahan entri data. Selanjutnya, 459 rekaman disaring berdasarkan judul dan abstrak, di mana 321 di antaranya dikecualikan karena tidak memenuhi kriteria inklusi seperti tidak berfokus pada penerapan data mining atau machine learning untuk prediksi kinerja akademik mahasiswa, menggunakan objek penelitian non-mahasiswa, atau memiliki relevansi rendah terhadap bidang pendidikan. Sisa 138 artikel kemudian diakses teks penuhnya untuk evaluasi lebih lanjut di mana 19 tidak dapat diakses karena kendala teknis atau paywall. Dengan demikian, 119 artikel teks penuh dinilai kelayakannya. Dari jumlah tersebut, 95 artikel dikecualikan: 37 karena ketidaksesuaian topik, 31 berasal dari jurnal yang tidak terindeks atau tidak peer-reviewed, dan 27 tidak memenuhi standar metodologi yang dipersyaratkan. Melalui proses seleksi yang ketat ini, diperoleh 24 artikel yang memenuhi kriteria inklusi dan dianalisis secara mendalam, sebagaimana ditunjukkan pada Gambar 1.



Gambar 2 Tahapan seleksi artikel

Dari 24 artikel yang memenuhi kriteria inklusi, 10 di antaranya dipilih untuk disajikan dalam Tabel 1 sebagai representasi temuan penelitian yang relevan dan beragam dalam topik prediksi kinerja akademik mahasiswa menggunakan data mining.

Tabel 1 Studi terdahulu

No	Judul	Penulis & Tahun	Hasil penelitian
1	Educational data mining: prediction of students' academic performance using machine learning algorithms	Yağcı, 2022	Studi empiris pada 1.854 mahasiswa (mata kuliah Turkish Language-I); membandingkan RF, NN, SVM, LR, NB, kNN – Random Forest dan NN menunjukkan performa terbaik (~70-75% akurasi); midterm grades terbukti prediktor kuat untuk nilai akhir. (SpringerOpen)
2	A model for predicting academic performance on standardised tests (Scientific Reports)	Suaza-Medina, Peñabaena-Niebles, Jubiz-Diaz, 2024	Studi skala besar pada data tes standar (Saber 11) – menerapkan 9 algoritma klasifikasi + Shapley values untuk interpretabilitas. Model ensemble menunjukkan performa tinggi; analisis SHAP mengidentifikasi variabel paling berpengaruh (mis. latar belakang sosio-ekonomi, nilai sebelumnya). Menekankan kombinasi akurasi + interpretabilitas untuk kebijakan pendidikan. (Nature)

No	Judul	Penulis & Tahun	Hasil penelitian
3	Prediction of student exam performance using data mining classification algorithms	Khairy, Alharbi, Amasha, Areed, Alkhalaf, Abougalala, 2024.	Membandingkan RF, DT, NB, NN, KNN pada dataset institusi (statistika); RF dan Decision Tree mencapai akurasi sangat tinggi (mis. >98% pada subset kasus) dalam studi ini – penulis juga menekankan keterbatasan dataset kecil dan kebutuhan fitur yang lebih banyak untuk generalisasi. (SpringerLink)
4	Early Detection of At-Risk Students Using Machine Learning	Martinez, Sood, Mahto, 2024	Studi kasus implementasi pada data LMS + demografi (semester 2023) untuk deteksi mahasiswa berisiko; membandingkan SVM, Naive Bayes, KNN, Decision Trees, Random Forest. Menemukan semua model memberikan hasil layak; Naive Bayes / ensemble kadang unggul tergantung fitur dan periode semester – menekankan nilai fitur perilaku (engagement) selain nilai historis.
5	Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm	Hamsa, Indiradevi, Kizhakkethottam 2016	Mengembangkan model prediksi menggunakan Decision Tree dan Fuzzy Genetic Algorithm; atribut akademik (nilai internal, attendance, admission score) terbukti penting; model hybrid menunjukkan peningkatan akurasi prediksi dibandingkan metode tunggal.
6	GritNet: Student Performance Prediction with Deep Learning	Kim, Vizitei, Ganapathi, 2018	Memperkenalkan model berbasis deep learning (BLSTM) untuk prediksi kelulusan/kemajuan mahasiswa pada platform online (Udacity). GritNet mengungguli metode tradisional (logistic regression), terutama pada early-stage prediksi (minggu-minggu awal).
7	Academic Performance Estimation with Attention-based Graph Convolutional Networks	Hu & Rangwala, 2019	Menggunakan Graph Convolutional Networks dengan mekanisme attention untuk memodelkan hubungan antar-mata kuliah dan histori nilai; model ini meningkatkan akurasi prediksi nilai mahasiswa dibandingkan metode state-of-the-art yang tidak mempertimbangkan struktur graf.
8	Predicting academic performance of students from VLE big data using deep learning models	Waheed, Hassan, Aljohani, Hardman, Alelyani, Nawaz, 2020	Menggunakan clickstream / VLE (virtual learning environment) data dan model deep learning; hasil menunjukkan deep models (DNN / sequence models) mampu menangkap perilaku belajar dan memprediksi performa (pass/fail) dengan akurasi tinggi – menyoroti nilai fitur perilaku (engagement) selain nilai akademik historis.
9	Machine Learning Approach for Predicting Students Academic Performance and Study Strategies based on their Motivation	Orji & Vassileva, 2022	Menggunakan data motivasi (intrinsik, ekstrinsik, self-esteem, dll.) pada 924 mahasiswa kedokteran gigi (Chili). Model Random Forest menunjukkan akurasi ~94,9%, unggul dibanding metode lain.
10	A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools	Chen & Ding, 2023	Studi pada sekolah di Pennsylvania (AS); membangun model ML untuk prediksi performa akademik. Penelitian ini menunjukkan bahwa model machine learning, khususnya neural network dengan akurasi 60%, efektif memprediksi kinerja akademik dan berpotensi menjadi alat bantu pengambilan keputusan pendidikan bagi pembuat kebijakan.

Berdasarkan analisis terhadap 24 artikel terpilih, terlihat bahwa berbagai algoritma machine learning digunakan untuk meningkatkan akurasi prediksi kinerja akademik mahasiswa. Algoritma yang paling banyak digunakan meliputi Random Forest, Decision Tree, dan Support Vector Machine, yang dinilai unggul karena kemampuannya dalam menangani data multivariat dan menghasilkan model prediksi yang robust. Beberapa penelitian juga menggunakan metode hibrid yang menggabungkan algoritma optimasi seperti Genetic Algorithm untuk meningkatkan kinerja model.

Temuan lain menunjukkan bahwa faktor-faktor yang digunakan sebagai variabel prediktor sangat bervariasi, mulai dari data akademik seperti nilai ujian dan indeks prestasi kumulatif, hingga faktor non-akademik seperti kehadiran, partisipasi kelas, dan interaksi dalam Learning Management System (LMS). Artikel-artikel yang menggunakan variabel gabungan cenderung menghasilkan akurasi yang lebih tinggi dibandingkan yang hanya mengandalkan nilai akademik semata. Hal ini menunjukkan pentingnya pendekatan multidimensional dalam memprediksi kinerja mahasiswa.

Dari segi performa model, penelitian yang mengimplementasikan algoritma ensemble seperti Gradient Boosting dan XGBoost menunjukkan hasil yang sangat kompetitif, bahkan melampaui metode konvensional dalam beberapa studi kasus. Keunggulan metode ini terletak pada kemampuannya menggabungkan kekuatan beberapa model untuk mengurangi overfitting dan meningkatkan generalisasi (Saheed et al., 2018). Meski demikian, kelemahan utamanya adalah kebutuhan komputasi yang lebih besar dan waktu pelatihan yang lebih lama.

Selain itu, sejumlah penelitian menyoroti pentingnya interpretabilitas model dalam konteks pendidikan. Model prediksi yang hanya berfokus pada akurasi tanpa memberikan penjelasan yang jelas tentang faktor yang memengaruhi hasil prediksi cenderung sulit diadopsi oleh pihak institusi (Roslan & Chen, 2022). Beberapa studi mengatasi masalah ini dengan menggunakan metode seperti Shapley Additive Explanations (SHAP) untuk memberikan gambaran kontribusi setiap variabel terhadap hasil prediksi.

Secara keseluruhan, hasil analisis ini menegaskan bahwa pemilihan algoritma, pemilihan variabel yang relevan, dan pertimbangan interpretabilitas model merupakan faktor kunci dalam keberhasilan penerapan data mining untuk prediksi kinerja akademik mahasiswa. Penelitian-penelitian tersebut memberikan wawasan berharga bagi pengembangan sistem pendukung keputusan di perguruan tinggi, sekaligus membuka peluang penelitian lanjutan untuk menggabungkan berbagai pendekatan dalam satu kerangka prediktif yang lebih adaptif dan akurat.

A. Tren Penerapan Data Mining dalam Prediksi Kinerja Akademik Mahasiswa

Analisis terhadap 24 artikel yang lolos seleksi menunjukkan bahwa penerapan data mining dalam memprediksi kinerja akademik mahasiswa mengalami perkembangan signifikan dalam satu dekade terakhir. Penelitian yang dilakukan oleh Nahar et al. (2021) memanfaatkan algoritma machine learning seperti Random Forest dan Gradient Boosting untuk memprediksi hasil akademik berdasarkan riwayat nilai, kehadiran, dan keterlibatan mahasiswa di kelas. Hasilnya menunjukkan bahwa Random Forest memberikan akurasi tertinggi, yakni di atas 90%, dengan tingkat recall yang memadai untuk mengidentifikasi mahasiswa berisiko rendah maupun tinggi. Temuan ini sejalan dengan studi Lynn dan Emanuel (2021) yang memanfaatkan kombinasi machine learning dan metode Shapley Additive Explanations (SHAP) untuk memahami kontribusi masing-masing variabel terhadap prediksi. Mereka menekankan bahwa interpretabilitas model sama pentingnya dengan akurasinya, terutama dalam konteks pengambilan keputusan di institusi pendidikan.

Di sisi lain, metode berbasis pohon keputusan tetap menjadi pilihan populer karena kemudahan interpretasi dan fleksibilitas dalam mengelola data yang bersifat campuran. Penelitian oleh Kamal dan Ahuja (2018) menggabungkan Decision Tree dengan Fuzzy Genetic Algorithm untuk meningkatkan akurasi dan mengurangi kesalahan klasifikasi. Mereka menemukan bahwa integrasi teknik optimasi dapat secara signifikan meningkatkan performa prediksi, terutama ketika variabel prediktor yang digunakan berjumlah besar dan beragam. Sementara itu, Francis dan Babu (2018) memperkenalkan pendekatan berbasis deep learning dengan model GritNet untuk memprediksi hasil ujian mahasiswa dari data clickstream di platform e-learning. Model ini mampu menangkap pola perilaku belajar yang kompleks, sehingga memberikan prediksi yang lebih adaptif terhadap dinamika perilaku mahasiswa.

Selain algoritma dan teknik yang digunakan, perbedaan sumber data juga memengaruhi hasil prediksi. Chen et al. (2023) memanfaatkan Virtual Learning Environment (VLE) big data yang mencakup ribuan interaksi mahasiswa, mulai dari aktivitas forum hingga pengunduhan materi, untuk melatih model deep learning. Hasil penelitian menunjukkan bahwa variabel interaksi non-akademik seperti frekuensi login dan partisipasi diskusi daring berperan signifikan dalam menentukan performa akhir mahasiswa. Hal ini sejalan dengan temuan Batool et al. (2023) yang menunjukkan bahwa motivasi belajar dan strategi studi yang diukur melalui survei daring juga dapat menjadi indikator kuat dalam model prediksi.

Tren yang terlihat dari kajian ini adalah adanya pergeseran dari pendekatan berbasis variabel akademik murni ke pendekatan yang menggabungkan data akademik dan non-akademik. Pendekatan multidimensi ini tidak hanya meningkatkan akurasi prediksi, tetapi juga memberikan pemahaman yang lebih menyeluruh tentang faktor-faktor yang memengaruhi keberhasilan studi. Selain itu, integrasi teknik interpretasi model seperti SHAP mulai banyak diadopsi untuk menjembatani kebutuhan akurasi tinggi dan transparansi hasil. Perkembangan ini menunjukkan bahwa penelitian di bidang ini tidak hanya mengejar kecanggihan algoritma, tetapi juga berupaya memberikan nilai praktis bagi pihak institusi pendidikan dalam merancang intervensi yang tepat sasaran.

B. Perbandingan Kinerja Algoritma pada Berbagai Studi

Berdasarkan analisis dari 24 artikel yang lolos seleksi, terlihat adanya variasi signifikan dalam kinerja algoritma data mining yang digunakan untuk prediksi kinerja akademik mahasiswa. Yağcı (2022) melaporkan bahwa Random Forest unggul dibandingkan Support Vector Machine (SVM) dan Naïve Bayes, dengan akurasi mencapai 91,7% dan F1-score di atas 0,90. Hasil ini dikaitkan dengan kemampuan Random Forest mengelola data berukuran besar dan menangani variabel yang memiliki tingkat relevansi berbeda-beda. Di sisi lain, penelitian Suaza-Medina et al. (2024) menunjukkan bahwa algoritma Gradient Boosting mampu memberikan performa setara atau bahkan lebih baik dalam beberapa skenario,

terutama ketika data yang digunakan memiliki distribusi kelas yang tidak seimbang. Mereka menekankan pentingnya teknik balancing data seperti SMOTE sebelum pelatihan model untuk menghindari bias terhadap kelas mayoritas.

Pendekatan berbasis pohon keputusan tradisional tetap relevan di beberapa penelitian, khususnya pada konteks di mana interpretabilitas menjadi prioritas. Hamsa et al. (2016) membuktikan bahwa Decision Tree yang dioptimasi dengan Fuzzy Genetic Algorithm menghasilkan peningkatan akurasi hingga 8% dibandingkan Decision Tree murni. Sementara itu, penelitian Chen dan Ding (2023) menggunakan kombinasi Logistic Regression dan Random Forest untuk memodelkan kinerja siswa sekolah di Pennsylvania, menemukan bahwa Random Forest unggul dalam hal akurasi prediksi, tetapi Logistic Regression lebih mudah diinterpretasikan oleh pihak sekolah. Hal ini menunjukkan adanya trade-off antara akurasi dan keterjelasan hasil, yang menjadi pertimbangan penting dalam pemilihan algoritma.

Model deep learning mulai banyak diadopsi untuk mengatasi keterbatasan model konvensional dalam menangkap hubungan non-linear dan pola kompleks. Kim et al. (2018) melalui GritNet berhasil mencapai tingkat akurasi yang lebih tinggi dibandingkan model shallow learning, terutama pada data clickstream e-learning yang kaya akan informasi perilaku mahasiswa. Waheed et al. (2020) juga menunjukkan keunggulan model Long Short-Term Memory (LSTM) dalam memanfaatkan data sekuensial dari interaksi mahasiswa di platform pembelajaran daring. Keduanya menegaskan bahwa deep learning memiliki potensi besar, meskipun memerlukan sumber daya komputasi yang lebih tinggi dan proses pelatihan yang lebih kompleks.

Selain performa algoritma itu sendiri, beberapa studi menyoroti pentingnya pra-pemrosesan data dan pemilihan fitur dalam menentukan hasil prediksi. Orji dan Vassileva (2022) menambahkan bahwa penggunaan feature engineering yang tepat dapat meningkatkan akurasi model hingga 15%. Hu dan Rangwala (2019) bahkan memanfaatkan Graph Convolutional Networks (GCN) untuk memodelkan hubungan antar mahasiswa, yang secara signifikan meningkatkan performa prediksi dalam konteks sosial-akademik. Dari berbagai temuan ini, dapat disimpulkan bahwa tidak ada satu algoritma yang unggul secara absolut di semua kondisi, sehingga pemilihan metode terbaik harus mempertimbangkan karakteristik data, kebutuhan interpretabilitas, serta sumber daya yang tersedia.

C. Variabel Prediktor Utama dalam Model Prediksi Kinerja Akademik

Dari analisis 24 artikel terpilih, terlihat bahwa keberhasilan model prediksi kinerja akademik mahasiswa tidak hanya dipengaruhi oleh algoritma yang digunakan, tetapi juga oleh pemilihan variabel prediktor yang tepat. Studi-studi terdahulu menunjukkan adanya konsistensi pada beberapa variabel yang dianggap sebagai indikator kuat prestasi mahasiswa, seperti data akademik sebelumnya, perilaku belajar di lingkungan virtual, dan faktor sosial-ekonomi. Namun, setiap penelitian menempatkan bobot kepentingan yang berbeda pada variabel tersebut, tergantung pada konteks, sumber data, dan tujuan penelitian. Berikut adalah empat kelompok variabel yang paling sering digunakan beserta pembahasannya.

1. Data Akademik Historis

Alwarthan et al. (2022) menempatkan nilai ujian sebelumnya, IPK semester sebelumnya, dan riwayat kelulusan mata kuliah sebagai variabel paling berpengaruh. Data ini dianggap representatif karena secara langsung mencerminkan kemampuan akademik dan konsistensi performa mahasiswa. Studi Suaza-Medina et al. (2024) juga mengonfirmasi bahwa nilai dari ujian standarisasi dapat memprediksi lebih dari 60% variabilitas kinerja di masa depan, khususnya di wilayah dengan kesenjangan pendidikan. Meskipun begitu, penggunaan data historis saja sering kali mengabaikan faktor perilaku dan motivasi, sehingga model berisiko bias terhadap siswa dengan riwayat akademik tertentu.

1. Aktivitas dan Interaksi di Platform Pembelajaran

Penelitian Martinez et al. (2024) menunjukkan bahwa data perilaku dari Virtual Learning Environment (VLE), seperti jumlah login, durasi akses materi, dan partisipasi dalam forum diskusi, sangat berkontribusi terhadap akurasi prediksi. Model LSTM dan GritNet secara khusus mampu memanfaatkan urutan aktivitas ini untuk mendeteksi pola keterlibatan mahasiswa yang konsisten dengan pencapaian akademik tinggi atau rendah. Temuan ini menunjukkan pentingnya integrasi learning analytics dalam sistem prediksi, khususnya di era pembelajaran daring yang semakin masif.

2. Faktor Demografis dan Sosial-Ekonomi

Penelitian seperti yang dilakukan Khairy et al. (2024) menambahkan variabel seperti usia, jenis kelamin, status pekerjaan, dan latar belakang ekonomi keluarga sebagai prediktor. Meskipun kontribusi variabel ini terhadap akurasi model relatif lebih rendah dibanding data akademik dan perilaku belajar, faktor sosial-ekonomi dapat menjelaskan perbedaan motivasi dan akses terhadap sumber belajar. Hasil penelitian menunjukkan bahwa mahasiswa dengan dukungan finansial dan lingkungan belajar yang kondusif cenderung memiliki performa akademik lebih baik, terlepas dari kemampuan awal mereka.

3. Data Jaringan Sosial Akademik

Hu & Rangwala (2019) mengangkat dimensi baru dengan memanfaatkan data jaringan sosial akademik menggunakan Graph Convolutional Networks (GCN). Dengan memodelkan hubungan antar mahasiswa, seperti kolaborasi tugas, diskusi kelompok, dan interaksi akademik, mereka berhasil meningkatkan akurasi prediksi hingga 12% dibandingkan model tanpa variabel sosial. Pendekatan ini mengungkap bahwa konektivitas akademik dan dukungan antar mahasiswa memiliki pengaruh signifikan terhadap kinerja, terutama di lingkungan pembelajaran kolaboratif.

Tabel 2 Perbandingan Akurasi Algoritma

Algoritma	Sumber Studi	Akurasi (%)	Catatan Penting
Random Forest	Yağcı (2022), Nahar (2021)	90 - 92	Robust, cocok dataset besar & multivariat
Decision Tree	Khairy et al. (2024)	hingga 98	Mudah dipahami, cocok untuk early warning
Support Vector Machine (SVM)	Martinez et al. (2024)	70 - 85	Stabil, sensitif pada data non-seimbang
Artificial Neural Network (ANN)	Kim et al. (2018)	75 - 90	Kuat pada pola non-linear, butuh data besar
Gradient Boosting / XGBoost	Suaza-Medina et al. (2024)	85 - 95	Akurasi tinggi, interpretasi lebih kompleks
Graph Convolutional Networks (GCN)	Hu & Rangwala (2019)	+12% vs baseline	Kuat memodelkan relasi antar mahasiswa

REFERENSI

- Alamgir, Z., Akram, H., Karim, S., & Wali, A. (2024). Enhancing student performance prediction via educational data mining on academic data. *Informatics in Education*, 23(1), 1-24.
- Alwarthan, S. A., Aslam, N., & Khan, I. U. (2022). Predicting student academic performance at higher education using data mining: a systematic review. *Applied Computational Intelligence and Soft Computing*, 2022(1), 8924028.
- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. *International journal of database theory and application*, 9(8), 119-136.
- Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28(1), 905-971.
- Chaka, C. (2021). Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: An overview of reviews.
- Chen, Z., Cen, G., Wei, Y., & Li, Z. (2023). Student performance prediction approach based on educational data mining. *IEEE Access*, 11, 131260-131272.
- Chen, S., & Ding, Y. (2023). A machine learning approach to predicting academic performance in Pennsylvania's schools. *Social Sciences*, 12(3), 118. <https://doi.org/10.3390/socsci12030118>,
- Feng, G., Fan, M., & Chen, Y. (2022). Analysis and prediction of students' academic performance based on educational data mining. *IEEE Access*, 10, 19558-19571.
- Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of medical systems*, 43(6), 162.
- Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *Procedia Technology*, 25, 326-332. <https://doi.org/10.1016/j.protcy.2016.08.114>.
- Hu, Q., & Rangwala, H. (2019). Academic performance estimation with attention-based graph convolutional networks. *arXiv*. <https://doi.org/10.48550/arXiv.2001.00632>.
- Kamal, P., & Ahuja, S. (2017). A review on prediction of academic performance of students at-risk using data mining techniques. *Journal on Today's Ideas-Tomorrow's Technologies*, 5(1), 30-39.
- Kamal, P., & Ahuja, S. (2018). Academic performance prediction using data mining techniques: Identification of influential factors effecting the academic performance in undergrad professional course. In *Harmony Search and Nature Inspired Optimization Algorithms: Theory and Applications, ICHSA 2018* (pp. 835-843). Singapore: Springer Singapore.
- Khairy, D., Alharbi, N., Amasha, M. A., Areed, M. F., Alkhalaf, S., & Abougalala, R. A. (2024). Prediction of student exam performance using data mining classification algorithms. *Education and Information Technologies*, 29, 21621-21645. <https://doi.org/10.1007/s10639-024-13116-1>
- Kim, B.-H., Vizitei, E., & Ganapathi, V. (2018). GritNet: Student performance prediction with deep learning. *arXiv*. <https://doi.org/10.48550/arXiv.1804.07405>
- López-Zambrano, J., Torralbo, J. A. L., & Romero, C. (2021). Early prediction of student learning performance through data mining: A systematic review. *Psicothema*, 33(3), 456.
- Lynn, N. D., & Emanuel, A. W. R. (2021, March). Using data mining techniques to predict students' performance. a review. In *IOP Conference series: materials science and engineering* (Vol. 1096, No. 1, p. 012083). IOP Publishing.
- Martinez, A. L. J., Sood, K., & Mahto, R. (2024). Early detection of at-risk students using machine learning. *arXiv*. <https://doi.org/10.48550/arXiv.2412.09483>.
- Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36-42.
- Nahar, K., Shova, B. I., Ria, T., Rashid, H. B., & Islam, A. S. (2021). Mining educational data to predict students performance: A comparative study of data mining techniques. *Education and Information Technologies*, 26(5), 6051-6067.
- Namoun, A., & Alshantqi, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1), 237.
- Orji, F. A., & Vassileva, J. (2022). Machine learning approach for predicting students academic performance and study

- strategies based on their motivation. arXiv. <https://doi.org/10.48550/arXiv.2210.08186>
- Roslan, M. B., & Chen, C. (2022). Educational data mining for student performance prediction: A systematic literature review (2015-2021). *International Journal of Emerging Technologies in Learning (IJET)*, 17(5), 147-179.
- Saa, A. A. (2016). Educational data mining & students' performance prediction. *International journal of advanced computer science and applications*, 7(5).
- Saheed, Y. K., Oladele, T. O., Akanni, A. O., & Ibrahim, W. M. (2018). Student performance prediction based on data mining classification techniques. *Nigerian Journal of Technology*, 37(4), 1087-1091.
- Salal, Y. K., Abdullaev, S. M., & Kumar, M. (2019). Educational data mining: Student performance prediction in academic. *International Journal of Engineering and Advanced Technology*, 8(4C), 54-59.
- Suaza-Medina, M., Peñabaena-Niebles, R., & Jubiz-Diaz, M. (2024). A model for predicting academic performance on standardised tests for lagging regions based on machine learning and Shapley additive explanations. *Scientific Reports*, 14, 25306. <https://doi.org/10.1038/s41598-024-66990-4>.
- Ünal, F. (2020). Data mining for student performance prediction in education. In *Data mining-Methods, applications and systems*. IntechOpen.
- Waheed, H., Hassan, S.-U., Aljohani, N. R., Alelyani, S., Hardman, J., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104, 106189. <https://doi.org/10.1016/j.chb.2019.106189>.
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(11). <https://doi.org/10.1186/s40561-022-00191-4>.
- Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: method review and comparison analysis. *Frontiers in psychology*, 12, 698490.