



Analisis Perbandingan Algoritma Machine Learning Dalam Prediksi Akademik Mahasiswa Berbasis Data

¹Apriliana Putri Maulikha, ²Najwa Aisha Hirania*, ³Fatecha Athallah Ahmad, ⁴Celvin Rasya Pamungkas, ⁵Muhammad Faiz Ardiansyah, ⁶Faridhatun Nikmah

^{1,2,3,4,5,6}Program Studi Sains Data, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Raden Mas Said Surakarta
Email: 257411046@mhs.uinsaid.ac.id¹, 257411047@mhs.uinsaid.ac.id², 257411036@mhs.uinsaid.ac.id³, 257411033@mhs.uinsaid.ac.id⁴, 257411051@mhs.uinsaid.ac.id⁵, faridhatunnikmah28@gmail.com⁶

*Corresponding author: Najwa Aisha Hirania

ABSTRAK

Tingginya angka kegagalan akademik merupakan salah satu indikator signifikan yang memengaruhi kualitas dan akreditasi institusi perguruan tinggi. Penelitian ini bertujuan untuk menganalisis perbandingan algoritma *Machine Learning* dalam memprediksi performa akademik dan merumuskan strategi cara mengatasi mahasiswa yang teridentifikasi berisiko gagal. Metode penelitian yang digunakan yaitu *Systematic Literature Review* (SLR) terhadap literatur yang dipublikasikan antara tahun 2019-2025. Hasil analisis menunjukkan bahwa algoritma *Decision Tree* (C4.5) merupakan model paling optimal untuk pengelompokan risiko dengan tingkat akurasi mencapai hingga 97.80% yang mampu menangani dalam data administratif yang bersifat kategorikal. Sementara itu, analisis regresi linear menunjukkan adanya korelasi linear yang kuat antara perilaku belajar mahasiswa dengan capaian akademik, variabel presensi juga memberikan kontribusi sebesar 94% ($R^2 = 0,94$) terhadap akurasi prediksi. Tidak hanya itu temuan lain mengungkapkan bahwa integrasi data *real time* dari *Learning Management System* (LMS) secara signifikan dapat meningkatkan presisi pada model dibandingkan dengan penggunaan data historis statis. Simpulan penelitian ini menegaskan bahwa penerapan hasil prediksi melalui *Early Warning System* yang diintegrasikan dengan bimbingan akademik intensif oleh Dosen Pembimbing Akademik (DPA) merupakan solusi efektif untuk menekan angka kegagalan studi secara dini dan objektif.

Kata Kunci: *Decision Tree, Early Warning System, Keberhasilan Akademik, Machine Learning, System Literature Review*

ABSTRACT

The high rate of academic failure is a significant indicator that affects the quality and accreditation of higher education institutions. This research aims to analyze the comparison of Machine Learning algorithms in predicting academic performance and formulate strategies for dealing with students identified as being at risk of failure. The research method used is Systematic Literature Review (SLR) of literature published between 2019-2025. The analysis results show that the Decision Tree algorithm (C4.5) is the most optimal model for risk grouping with an accuracy level of up to 97.80% which is able to handle categorical administrative data. Meanwhile, linear regression analysis shows that there is a strong linear correlation between student learning behavior and academic achievement, the presence variable also contributes 94% ($R^2 = 0.94$) to prediction accuracy. Not only that, other findings reveal that the integration of real time data from the Learning Management System (LMS) can significantly increase model precision compared to the use of static historical data. The conclusion of this research confirms that the application of predicted results through the Early Warning System which is integrated with intensive academic guidance by Academic Supervisors (DPA) is an effective solution for reducing the number of study failures early and objectively.

Keywords: *Academic Success, Decision Tree, Early Warning System, Machine Learning, System Literature Review*

1. PENDAHULUAN

Perkembangan transformasi digital saat ini membuat sektor perguruan tinggi telah menjadi produsen data yang sangat besar. Setiap aktivitas mahasiswa, mulai dari proses pendaftaran hingga interaksi di dalam *Learning Management System* (LMS) meninggalkan jejak digital yang berharga. Namun, fenomena yang sering terjadi di Indonesia adalah “data gelap” (*dark data*) di mana tumpukan data tersebut hanya berakhir sebagai arsip administratif tanpa diolah secara strategis untuk mendukung kesuksesan mahasiswa (Romero & Ventura, 2020). Fenomena *dark data* ini sejalan dengan pandangan Baker (2021) yang menyatakan bahwa tantangan utama di institusi pendidikan pada abad ke-21 bukanlah kekurangan data, melainkan ketidakmampuan institusi dalam mengkolaborasikan kecerdasan buatan (*Artificial Intelligence*) untuk menghasilkan kebijakan yang tepat sasaran.

Di Indonesia, tantangan ini semakin nyata adanya desakan akreditasi dari Ban Akreditasi Nasional Perguruan Tinggi (BAN-PT) yang menuntut dalam standar efisiensi lulusan yang tinggi. Menurut Arifin dkk. (2022) menyatakan bahwa kegagalan akademik yang tidak terdeteksi sejak dini ini membuat efek yang merugikan mulai dari penurunan produktivitas institusi hingga sampai pada beban psikologis mahasiswa. Dengan adanya itu semua kualitas sebuah universitas seringkali dinilai dari efisiensi lulusannya. Masalah mendasar yang ditemukan adalah pada sistem evaluasi konvensional yang cenderung bersifat periodik, sehingga tindakan bantuan terlambat dilakukan karena penurunan performa baru disadari saat nilai akhir keluar.

Sains data hadir sebagai solusi dengan menawarkan kemampuan prediktif melalui algoritma Machine Learning yang berbeda dengan statistik konvensional. Machine Learning mampu menangkap pola perilaku belajar yang kompleks dari ribuan data historis mahasiswa untuk memproyeksikan performa akademik mereka di masa depan (Alam et al., 2019). Dengan akurasi tinggi, institusi dapat memetakan mahasiswa ke dalam kategori risiko tertentu sebelum masa studi berakhir. Beberapa penelitian di Indonesia telah membuktikan efektivitas teknologi ini. Sebagai contoh, penggunaan algoritma Decision Tree (C4.5) dalam lingkungan akademik didukung oleh Fatoni dkk. (2026) yang menyatakan bahwa model ini sangat efektif karena mampu menghasilkan aturan keputusan (*decision rules*) yang mudah dipahami oleh staf pengajar non-teknis.

Pendekatan komparatif sangat diperlukan karena setiap algoritma memiliki karakteristik. Menurut Suwartini dkk. (2024) menyatakan bahwa meskipun model seperti Naive Bayes memiliki kecepatan dengan proses yang tinggi namun, model tersebut membutuhkan variabel eksternal lain, seperti latar belakang ekonomi, untuk mendekati akurasi model dengan berbasis pohon keputusan. Sementara itu, model regresi linear menunjukkan bahwa variabel seperti kehadiran dan tugas memiliki kontribusi signifikan sebesar 0,94 atau 94% terhadap pembentukan nilai akhir (Salsabila et al., 2024).

Pemanfaatan data sangat penting karena berkaitan dengan pergeseran gaya belajar mahasiswa di tengah perkembangan digital saat ini. Okubo (2017) menunjukkan bahwa jejak digital mahasiswa pada Learning Management System (LMS) merupakan cerminan dari kedisiplinan diri. Mahasiswa yang jarang mengakses materi digital atau terlambat mengumpulkan tugas pada minggu awal semester memiliki korelasi linear terhadap potensi kegagalan pada akhir semester. Hal ini didukung dengan temuan Education & Studies (2024) dalam tinjauan sistematis yang menunjukkan bahwa algoritma Machine Learning mampu mendeteksi penurunan motivasi mahasiswa jauh lebih cepat dibandingkan dengan pengamatan langsung oleh dosen. Tanpa adanya sistem otomatis dalam memproses data ini intervensi yang dilakukan institusi pendidikan akan terlambat atau reaktif dalam menangani permasalahan ini (Shafiq et al., 2022).

Meskipun penelitian-penelitian tersebut telah berhasil meraih tingkat akurasi yang tinggi, terdapat kesenjangan penelitian (*research gap*) di mana sebagian besar studi berhenti pada tahap pelaporan nilai hasil di atas kertas (nilai akurasi algoritma). Kekosongan pembahasan mengenai hasil prediksi masih ada, sehingga belum diintegrasikan ke dalam langkah-langkah intervensi nyata dan sistematis di institusi perguruan tinggi. Protokol konkret bagi dosen pembimbing untuk menangani mahasiswa yang telah terdeteksi berisiko gagal oleh sistem masih belum dirumuskan.

Oleh karena itu, penelitian ini memfokuskan pada strategi penanganan mahasiswa yang teridentifikasi berisiko gagal melalui pemanfaatan hasil perbandingan algoritma Machine Learning. Alasan mendasar dilakukannya penelitian lanjutan ini adalah untuk memastikan bahwa kemajuan teknologi prediksi tidak hanya menjadi laporan statistik saja tetapi dapat menjadi instrumen penyelamatan akademik yang fungsional. Hal ini sangat penting mengingat angka drop-out disebabkan oleh keterlambatan respon institusi terhadap penurunan angka performa mahasiswa yang sebenarnya sudah dapat terbaca oleh data.

Dengan demikian, penelitian ini hadir untuk menjembatani antara kompleksitas teknik Sains Data dengan kebutuhan praktis dunia pendidikan, serta diharapkan dapat memberikan kontribusi yang nyata bagi pengembangan Standar Operasional Prosedur (SOP) bimbingan akademik yang berbasis nyata (evidence-based) di lingkungan perguruan tinggi.

2. METODE PENELITIAN

Penelitian ini menggunakan metode deskriptif dengan pendekatan Systematic Literature Review (SLR). Menurut **Kitchenham & Charters (2007)** SLR merupakan metode penelitian yang bertujuan untuk mengidentifikasi, mengevaluasi, dan menginterpretasikan seluruh bukti penelitian yang relevan terhadap pertanyaan penelitian atau fenomena tertentu secara sistematis. Pendekatan ini dipilih untuk mengintegrasikan berbagai temuan berbasis eksperimen dari penelitian terdahulu mengenai efektivitas algoritma *Machine Learning* di institusi pendidikan tinggi.

Penggunaan Systematic Literature Review (SLR) memungkinkan peneliti untuk menarik kesimpulan yang lebih luas, objektif, dan kredibel dibandingkan hanya mengandalkan studi kasus tunggal (Miles & Huberman, 1994). Melalui metode ini, dilakukan perbandingan performa antar algoritma secara sistematis guna menemukan pola penanganan terbaik bagi mahasiswa yang berisiko gagal. Data yang digunakan merupakan data sekunder dari jurnal ilmiah terakreditasi melalui basis data Google Scholar, Garuda (Garba Rujukan Digital), dan ResearchGate.

Dalam pelaksanaannya, peneliti menetapkan kriteria inklusi dan eksklusi yang ketat berdasarkan protokol SLR untuk menjamin kualitas serta validitas data yang disintesis (Kitchenham & Charters, 2007). Kriteria tersebut mencakup artikel yang dipublikasikan dalam rentang tahun 2019—2025, memiliki fokus penelitian pada prediksi performa akademik mahasiswa di tingkat perguruan tinggi, menerapkan algoritma *Machine Learning*, serta menyajikan indikator akurasi yang jelas seperti Accuracy, Precision, Recall, atau R-squared. Proses pencarian dilakukan menggunakan kata kunci seperti “Machine Learning”, “Prediksi Akademik”, dan “Mahasiswa”. Instrumen peneliti yang digunakan adalah lembar ekstraksi data (data extraction sheet) untuk mencatat informasi kata kunci dari setiap literatur yang terpilih secara terorganisir.

Proses analisis data sintesis dalam penelitian dilakukan melalui proses interaktif berkelanjutan yang merujuk pada model Miles, Huberman, dan Saldana (2014), Teknik ini terdiri dari tiga tahapan utama yang saling berkaitan:

- a. **Kondensasi Data**, peneliti melakukan filtrasi terhadap variabel-variabel yang dianggap non-relevan dan memfokuskan analisis pada variabel kunci seperti Indeks Prestasi (IPK), tingkat kehadiran, dan aktivitas digital mahasiswa. Pada tahap ini, ukuran performa algoritma yang digunakan seperti R-squared (Koefisien determinasi) dan nilai akurasi (R^2 , Akurasi) dikumpulkan secara sistematis dari literatur sumber.
- b. **Penyajian Data (*Data Display*)**, Peneliti menyusun seluruh data teknis yang telah dikondensasi ke dalam matriks perbandingan. Penyajian data dalam bentuk tabel yang bertujuan untuk mempermudah identifikasi pola keunggulan, kelemahan, dan karakteristik unik dari masing-masing algoritma yang diteliti.
- c. **Penarikan Kesimpulan dan Verifikasi**, Peneliti merumuskan strategi konkret mengenai cara mengatasi mahasiswa yang teridentifikasi berisiko gagal berdasarkan pola data yang ditemukan. Hasil rumusan ini kemudian diverifikasi kembali menggunakan teori-teori pendukung dan hasil penelitian terdahulu yang relevan untuk memastikan validitas rekomendasi yang diberikan.

Berdasarkan protokol *Systematic Literature Review* (SLR) yang telah dijelaskan, berikut adalah hasil dari sintesis data dari berbagai literatur terpilih. Bagian ini akan menguraikan perbandingan performa metrik antar algoritma dan perancangan kerangka kerja intervensi pada akademik yang dapat diimplementasikan oleh institusi pendidikan perguruan tinggi.

3. HASIL DAN PEMBAHASAN

HASIL

Berdasarkan proses Systematic Literature Review (SLR) yang dilakukan terhadap empat literatur utama, peneliti mengekstraksi data performa algoritma yang digunakan untuk memprediksi performa akademik mahasiswa. Hasil ekstraksi tersebut disajikan dalam tabel matriks perbandingan dibawah ini:

Tabel 1. Matriks Perbandingan Performa Algoritma dari Literatur Referensi

Penulis (Tahun)	Algoritma Utama	Metrik Performa	Variabel Paling Berpengaruh
Maulana dkk. (2024)	Decision Tree (C4.5)	Akurasi : 97,80%	IP Semester Awal, Nilai, UN
Salsabila dkk. (2024)	Regresi Linear	R ² : 94%	Presensi, Nilai, Tugas
Wicaksono dkk. (2019)	Naive Bayes	Akurasi : 82,15%	Latar Belakang Ekonomi, IPK
Andika dkk. (2025)	Deep Learning	Akurasi : Tinggi	Log Aktivitas Digital (LMS)

Berdasarkan Tabel 1, data menunjukkan bahwa algoritma Decision Tree memiliki tingkat akurasi tertinggi dalam mengklasifikasikan mahasiswa ke dalam kategori lulus atau gagal karena dapat memecahkan data kompleks menjadi aturan keputusan yang sederhana (seperti "Jika IPK < X dan Kehadiran < Y, maka Berisiko"). Sementara itu, Regresi Linear memberikan pemahaman yang sangat kuat bahkan mencapai 94% mengenai bagaimana variabel dari perilaku seperti kehadiran secara langsung membentuk nilai akhir mahasiswa.

PEMBAHASAN

Setelah dilakukan analisis mendalam terhadap keempat literatur, ditemukan perbedaan fokus yang signifikan di antara literatur tersebut. Maulana et al. (2024) membuktikan bahwa data statis, seperti nilai Ujian Nasional (UN) dan IP semester awal, memiliki kekuatan prediksi yang cukup stabil pada algoritma Decision Tree. Hal tersebut menunjukkan bahwa algoritma Decision Tree (C4.5) menjadi model yang paling akurat dalam mengelompokkan mahasiswa yang berisiko dengan persentase mencapai hingga 97,80%. Keunggulan ini dikarenakan algoritma mampu dalam memecahkan struktur data administratif yang kompleks menjadi aturan keputusan (decision rules) yang logis. Di institusi perguruan tinggi ini sangat krusial karena civitas akademika seperti Dosen Pembimbing Akademik (DPA) memerlukan penjelasan mengenai "mengapa" mahasiswa dikategorikan berisiko dan bukan hanya dari hasil prediksi tanpa dasar yang jelas.

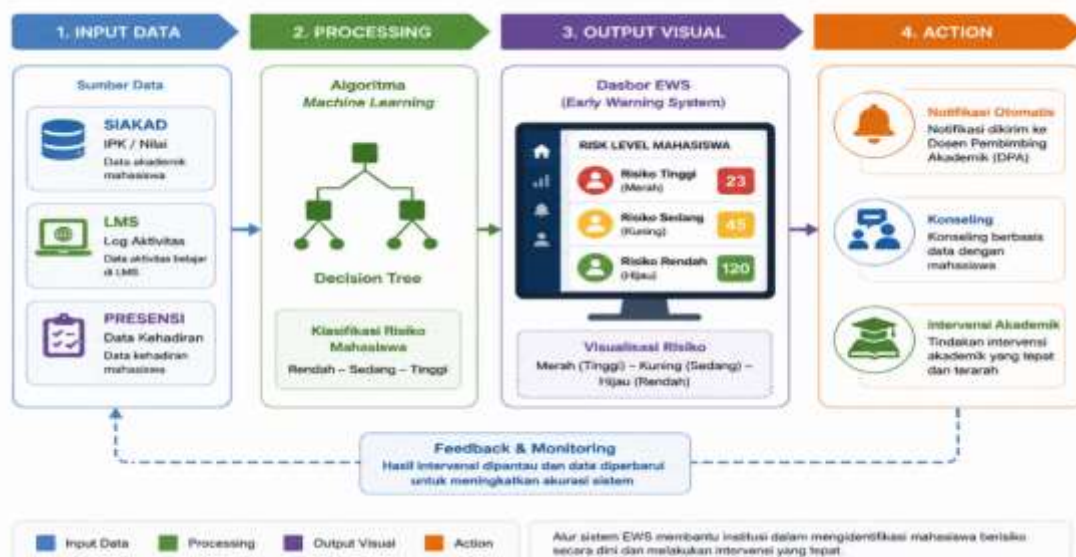
Di sisi lain, penggunaan regresi linear, menurut Salsabila dkk. (2024), memberikan pandangan yang berbeda namun signifikan. Nilai koefisien determinasi (R²) sebesar 0,94 menunjukkan bahwa variabel perilaku secara real-time seperti presensi mahasiswa, justru memiliki korelasi yang sangat kuat dalam memprediksi keberhasilan dan pembentukan nilai akhir. Temuan ini menegaskan bahwa tingkat kehadiran bukan hanya syarat untuk administrasi saja melainkan sebuah indikator utama kedisiplinan dan komitmen belajar mahasiswa yang dibaca secara numerik.

Sementara itu, Wicaksono dkk. (2019) memberikan perspektif penting dengan memasukkan variabel latar belakang ekonomi e salam permodelan. Temuan mereka menunjukkan bahwa faktor eksternal turut berperan penting terhadap keberlanjutan studi, meskipun akurasi algoritma Naive Bayes yang dihasilkan adalah sebesar 82,15%. Namun demikian, terdapat catatan penting terkait validitas metodologis yang menjadi keterbatasan dalam studi peninjauan literatur ini (disclaimer). Perbandingan performa metrik akurasi antara algoritma Decision Tree (97,80%) pada studi Maulana dkk. (2024) dan Naive Bayes (82,15%) pada studi Wicaksono dkk. (2019) tidak dapat diinterpretasikan secara langsung sebagai perbandingan yang setara (apple-to-apple). Secara esensi Sains Data, variabelitas nilai performa tersebut sangat dipengaruhi oleh perbedaan karakteristik dataset dasar yang digunakan masing-masing penelitian asli, termasuk ukuran sampel, distribusi kelas, teknik pra-pemrosesan data (preprocessing), serta variasi fitur masukan (input features). Oleh karena itu, pengutipan metrik akurasi dalam penelitian ini diposisikan

sebagai cerminan efektivitas algoritma pada batasan ruang lingkup studi kasus aslinya, bukan sebagai pembuktian mutlak bahwa satu algoritma secara universal lebih unggul dibanding algoritma lainnya.

Di samping itu, Andika dkk. (2025) membawa pembaruan dengan memanfaatkan Deep Learning untuk membaca log digital mahasiswa guna mengungkapkan perbedaan esensial antara pengguna data statis da dinamis. Data masa lalu, seperti nilai ujian masuk, memang memberikan gambaran kapasitas akademik mendasar. Namun, data perilaku secara real-time dari Learning Management System (LMS) terbukti mampu mendeteksi penurunan motivasi mahasiswa jauh lebih cepat dibandingkan dengan pengamatan manual. Oleh karena itu, integrasi antara data akademik statis maupun data perilaku dinamis merupakan kunci utama untuk mendapatkan akurasi prediksi yang paling komprehensif di institusi perguruan tinggi.

Penerapan *Machine Learning* di institusi perguruan tinggi bukan sekadar alat komputasi atau tren teknologi, melainkan sebuah instrumen strategis dalam pengelolaan keberhasilan mahasiswa. Hasil prediksi dari algoritma-algoritma di atas dapat diintegrasikan ke dalam sebuah *Early Warning System* (EWS). Sistem ini bekerja dengan cara mengolah data dinamis (contoh: kehadiran minggu ke-1 hingga ke-7) dan data statis seperti nilai masuk, untuk memberikan sebuah sinyal peringatan awal kepada pihak kampus sebelum masa ujian akhir tiba. Beberapa tahapan implementasi yang dapat dilakukan institusi perguruan tinggi:



Gambar 1. Alur Sistem *Early Warning System* (EWS)

a. Integrasi Data Terpusat (*Data Warehousing*)

Langkah pertama dalam penerapan ini adalah mengintegrasikan data dari berbagai sumber, seperti Sistem Informasi Akademik (SIAKAD), presensi sidik jari atau *QR Code*, hingga log aktivitas pada *Learning Management System* (LMS). Dengan mengumpulkan data ini ke dalam satu database terpusat, algoritma *Machine Learning* dapat menarik pola yang lebih akurat mengenai perilaku mahasiswa secara menyeluruh, baik dari sisi administratif maupun akademik.

b. Pembangunan Dasbor Peringatan Dini (*Early Warning System Dashboard*)

Dasbor *Early Warning System* (EWS) adalah tampilan visual atau panel kendali digital yang berfungsi sebagai pusat pemantauan performa mahasiswa secara *real time* bagi Dosen Pembimbing Akademik (DPA) untuk melakukan intervensi yang dapat diamati. Dalam Sains Data penerapan *Early Warning System* (EWS) dalam penelitian ini bukan sekedar untuk alat pemantau namun, sebagai mekanisme deteksi dini yang bersifat proaktif. Dengan memanfaatkan pembelajaran mesin (*Machine Learning*) sistem akan mampu melakukan evaluasi berkelanjutan pada data perilaku harian mahasiswa. Seperti pada hasil dari prediksi algoritma *Decision Tree* yang memiliki akurasi tinggi dan dapat divisualisasikan ke dalam dasbor interaktif yang dapat diakses oleh jajaran dekanat, ketua program studi, dan dosen pembimbing akademik. Dasbor ini tidak hanya menampilkan angka tetapi memberikan label visual (misalnya warna merah untuk risiko tinggi, kuning untuk waspada, dan hijau untuk aman). Hal ini memudahkan pengambilan kebijakan

untuk mengidentifikasi mahasiswa mana yang memerlukan perhatian khusus tanpa harus membedah ribuan data secara manual.

c. Personalisasi Intervensi Akademik

Salah satu keunggulan penerapan *Machine Learning* adalah mampu dalam memberikan analisis faktor penyebab. Jika model Regresi menunjukkan bahwa rendahnya nilai seorang mahasiswa disebabkan oleh persentase kehadiran yang buruk (kontribusi 94%), maka pihak institusi Perguruan Tinggi dapat memberikan intervensi yang bersifat non-akademik seperti konseling masalah pribadi atau ekonomi. Sebaliknya, jika penyebabnya adalah rendahnya nilai tugas maka intervensi yang diberikan bersifat akademik seperti pemberian kelas matrikulasi atau tutor sebaya.

d. Optimasi Kurikulum dan Metode Pengajaran

Selain bagi mahasiswa, penerapan ini bermanfaat bagi dosen untuk mengevaluasi metode pengajaran. Jika algoritma mendeteksi pola kegagalan yang meningkat pada mata kuliah tertentu di minggu-minggu awal, institusi dapat segera mengevaluasi apakah materi tersebut terlalu sulit atau metode penyampaiannya yang perlu diperbaiki. Dengan demikian, *Machine Learning* berfungsi sebagai alat evaluasi berkelanjutan demi menjaga standar mutu pendidikan di perguruan tinggi.

Sesuai dengan fokus penelitian ini, hasil prediksi algoritma harus ditindak lanjuti dengan langkah-langkah implementasi yang nyata. Berikut beberapa langkah-langkah untuk mengatasi Mahasiswa yang berisiko gagal:

a. Pemberitahuan Otomatis (*Automated Notification*)

Langkah ini bertujuan untuk menciptakan kesadaran mengenai kondisi akademik dari mahasiswa, dengan cara setelah seorang mahasiswa diklasifikasikan algoritma (seperti algoritma *Decision Tree*) yang dimana jika mahasiswa tersebut sudah masuk dalam kategori "Berisiko Tinggi", maka dengan segera sistem secara otomatis mengirimkan notifikasi kepada Dosen Pembimbing Akademik (DPA) dan mahasiswa yang bersangkutan tersebut. Adanya keterlibatan DPA dalam tahap ini sangat dibutuhkan karena DPA berfungsi untuk menjembatani antara kebijakan institusi dan kondisi psikologis mahasiswa.

Menurut (Arifin dkk, 2022) menyatakan bahwa pendampingan oleh DPA dengan berbasis data dapat meningkatkan motivasi belajar mahasiswa dan membantu mengidentifikasi hambatan non akademik yang tidak terbaca oleh sistem. Selain itu, bimbingan intensif dengan DPA terbukti efektif dalam memberikan dukungan emosional dan administratif yang diperlukan untuk mencegah mahasiswa mengambil keputusan untuk berhenti kuliah (*drop out*) akibat dari akumulasi masalah akademik yang tidak tertangani sejak dini (Shafiq et al., 2022).

b. Konseling Akademik Berbasis Data

Mahasiswa yang berisiko mengalami kegagalan tersebut akan dipanggil oleh Dosen Pembimbing Akademiknya untuk melakukan diskusi dua arah. Berbeda dengan pembimbingan pada biasanya, bimbingan ini berbasis data. Jika Algoritma menunjukkan risiko disebabkan oleh rendahnya kehadiran, maka DPA akan berfokus pada pencarian solusi atas kendala presensi mahasiswa (seperti masalah kesehatan atau masalah transportasi).

c. Program Pendampingan Sebaya

Mahasiswa yang diprediksi berisiko gagal pada mata kuliah teknis dapat diarahkan untuk mengikuti kelompok belajar yang dibimbing oleh mahasiswa berprestasi, seperti contoh yang sudah diimplementasikan di beberapa kampus yaitu Tutor Sebaya atau mungkin sebutan lainnya. Langkah ini cukup efektif untuk menurunkan tekanan psikologis mahasiswa saat belajar atau menghadapi mata kuliah yang cukup sulit.

d. Penyesuaian Materi di Platform Digital

Beberapa mahasiswa pasti memiliki log aktivitas rendah pada *Learning Management System (LMS)* salah satunya (seperti temuan pada model *Deep Learning*), dosen dapat memberikan "materi jembatan" yang lebih sederhana atau video tutorial tambahan untuk memicu kembali keaktifan mahasiswa di ruang digital.

4. KESIMPULAN DAN SARAN

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah disajikan pada Tabel 1, penelitian ini menyimpulkan bahwa implementasi Machine Learning memiliki efektivitas yang sangat signifikan dalam memprediksi performa akademik mahasiswa di perguruan tinggi. Salah satu algoritma berhasil menjadi model yang paling akurat hingga persentasenya mencapai 97,80% itu dibuktikan oleh algoritma Decision Tree (C4.5). Temuan ini menegaskan bahwa kegagalan akademik bukanlah sebuah kejadian acak, melainkan ini termasuk pola perilaku yang dapat diidentifikasi secara dini melalui variabel kunci seperti IP semester awal dan juga tingkat kehadiran.

Adapun cara yang paling efektif untuk mengatasi mahasiswa yang berisiko gagal, yaitu dengan melalui transformasi data prediksi menjadi tindakan nyata melalui Early Warning System (EWS). Hal itu disimpulkan dari penelitian yang dilakukan. Strategi investasi sistematis yang dilakukan meliputi pemberitahuan otomatis, konseling berbasis data, program pendampingan sebaya, dan penyesuaian materi digital, beberapa hal tersebut merupakan kunci utama dalam menurunkan angka drop out. Dengan demikian, peran Sains Data di institusi pendidikan beralih dari sekadar instrumen administratif menjadi instrumen penyelamatan akademik yang proaktif dan berbasis data, atau bisa disebut dengan data driven.

SARAN

Bagi Institusi Perguruan Tinggi

Disarankan bagi institusi untuk mulai integrasi data terpusat (data warehousing) yang menghubungkan SIAKAD dan Learning Management System (LMS) secara real-time. Pembangunan dasbor peringatan dini harus menjadi prioritas kebijakan agar pihak dekanat dan program studi dapat melakukan pemantauan performa mahasiswa secara objektif dan cepat.

Bagi Dosen Pembimbing Akademik (DPA)

DPA hendaknya memanfaatkan hasil prediksi machine Learning sebagai kompas dalam memberikan bimbingan. Pendekatan kepada mahasiswa yang masuk dalam kategori berisiko tinggi harus dilakukan secara empatik dan solutif dengan fokus pada akar masalah yang ditemukan oleh algoritma, baik itu kendala akademik maupun non-akademik

Bagi Peneliti Selanjutnya

Penelitian masa depan diharapkan dapat memperluas variabel penelitian dengan memasukkan aspek psikometrik atau kesehatan mental mahasiswa, mengingat faktor psikologis juga memiliki pengaruh besar terhadap keberlanjutan studi namun seringkali belum terekam dalam administratif kampus. Beberapa rekomendasi yang dapat digunakan untuk penelitian selanjutnya:

- a. **Eksplorasi Komparatif Algoritma Ensemble dan Deep Learning:** Direkomendasikan bagi peneliti masa depan untuk mengeksperimenkan model *Ensemble Learning* (misalnya *Random Forest* atau *XGBoost*) serta teknik *Deep Learning* yang lebih kompleks. Pengujian ini sebaiknya memanfaatkan data riil universitas untuk memvalidasi performa metriknya terhadap model *Decision Tree* (C4.5).
- b. **Integrasi Variabel Non-Akademik Kontemporer:** Penelitian berikutnya diharapkan mampu memetakan faktor risiko mahasiswa secara lebih sensitif dengan menambahkan variabel non-akademik. Faktor-faktor seperti kondisi mental/psikologis mahasiswa, dinamika sosial di kampus, hingga status finansial orang tua perlu diolah secara dinamis sebagai input baru dalam pemodelan *Machine Learning*.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Program Studi Sains Data, Fakultas Sains dan Teknologi, Universitas Islam Negeri Raden Mas Said Surakarta yang telah memberikan dukungan fasilitas akademik serta ekosistem ilmiah yang kondusif selama proses penyusunan penelitian ini. Penulis juga menyampaikan apresiasi terdalam kepada Ibu Faridhatun Nikmah selaku dosen pembimbing yang telah memberikan arahan, saran, dan bimbingan yang sangat berharga demi kesempurnaan artikel ini. Tidak lupa, terima kasih kepada rekan-rekan kelompok atas diskusi interaktif dan kerja sama tim yang solid hingga naskah artikel ini dapat diselesaikan dengan baik.

REFERENSI

- Andika, A. W., dkk. (2025). Penggunaan Deep Learning untuk Memprediksi Kinerja Akademik dan Memberi Dukungan yang Tepat bagi Siswa. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, 4(1).
- Arifin, N. Y., dkk. (2022). *Analisis Perancangan Sistem Informasi*. Media Sains Indonesia: Bandung.
- Baker, R. S. (2021). Artificial intelligence in education: Bringing it all together. *Journal of Artificial Intelligence in Education*, 43–54.
- Dahlan, M. S. (2021). *Statistik untuk Kedokteran dan Kesehatan: Deskriptif, Bivariat, dan Multivariat*. Epidemiologi Indonesia: Jakarta.
- Education, A., & Studies, D. (2024). Self-efficacy: A Systematic Literature Review. *Education and Development Studies*, 1(1).
- Fatoni, D. S., Ramadhan, F., P, M. A. H., A, M. A., Studi, P., Informasi, T., & Teknik, F. (2026). IMPLEMENTASI ALGORITMA DECISION TREE C4.5 UNTUK PREDIKSI KELULUSAN MAHASISWA: STUDI EKSPERIMENTAL. *Jurnal Teknologi Informasi*, 10(8), 1–16.
- Hidayat, M. T., dkk. (2023). Implementasi Algoritma Decision Tree C4.5 untuk Prediksi Kelulusan Mahasiswa. *Jurnal Teknik Informatika*, 16(2).
- Maulana, S., Premana, A., & Irawan, B. (2024). Prediksi Prestasi Akademik Siswa Terbaik Menggunakan Algoritma Decision Tree Berbasis Data Historis. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1). <https://ejournal.itn.ac.id/index.php/jati/article/view/8276>
- Murtaza, M., dkk. (2022). Student Retention and Educational Data Mining: A Systematic Literature Review. *IEEE Access*, 10, 1-15.
- Nugroho, A. S., dkk. (2021). *Sains Data: Teori dan Aplikasi*. Penerbit Universitas Indonesia (UI Press): Jakarta.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3). <https://doi.org/10.1002/widm.1355>
- Salsabila, K., Maulidia, N. F., Auliya, S., Hafid, Z., & Shinta, A. (2024). Analisis Prediksi Nilai Akhir Mahasiswa Menggunakan Algoritma Regresi Linear Berbasis Machine Learning pada Program Studi Teknologi Informasi Universitas Bina Sarana Informatika. *Jurnal Ilmiah Komputer*, 4(3), 16657–16662.
- Shafiq, D. A., Marjani, M., Ahamed, R., Habeeb, A., & Asirvatham, D. (2022). Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review. *IEEE Access*, 10, 72480–72503. <https://doi.org/10.1109/ACCESS.2022.3188767>
- Suwartini, J., Sari, I., Umar, E., Momo, L. L., Informatika, P. T., Stella, U., Sumba, M., Karya, J., Tambolaka, K., & Daya, S. B. (2024). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naïve Bayes Dan Decision Tree Pada Universitas Stella Maris Sumba. *Jurnal Sains dan Teknologi*, 02(03), 362–368.
- Zhao, Y., dkk. (2023). Application of Machine Learning in Higher Education: A Systematic Review. *Sustainability*, 15(11).