

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Dunia pendidikan saat ini menghadapi era digital dengan tantangan yang besar, di mana tuntutan akan kualitas sumber daya manusia semakin meningkat. Pendidikan tidak hanya sekedar proses transfer ilmu, melainkan sebuah usaha untuk mengembangkan potensi unik setiap individu secara optimal. Mengacu pada tuntutan pendidikan modern, pendekatan pembelajaran konvensional yang bersifat seragam dinilai sudah tidak lagi efektif dalam mengakomodasi perbedaan kemampuan dan gaya belajar siswa yang beragam. Sejalan dengan pendapat dari Wahyuni & Kurniawan (2025), klasifikasi kemampuan akademik siswa menjadi faktor krusial bagi tenaga pendidikan untuk memberikan *treatment* pembelajaran yang tepat sasaran, terutama dalam mendukung implementasi kurikulum yang menuntut perkembangan pengajaran [1].

Perkembangan teknologi saat ini telah membuka peluang baru untuk mengatasi keterbatasan dalam personalisasi pembelajaran melalui konsep *adaptive learning*. Pemanfaatan *Data Mining* dan *Machine Learning* dalam sektor pendidikan, atau yang dikenal sebagai *Education data Mining* (EDM), berkembang pesat sebagai bidang riset yang mampu menyaring pengetahuan dari data besar di lingkungan pendidikan. Berdasarkan temuan Pal dkk. (2025), sistem rekomendasi berbasis kecerdasan buatan memiliki potensi besar untuk menyatakan sumber daya pembelajaran yang di personalisasi sesuai dengan preferensi dan kebutuhan spesifik peserta didik [2]. Hal ini diperkuat oleh pendapat Atalla dkk. (2023) yang menyatakan bahwa sistem rekomendasi cerdas dapat mengotomatisasi pemodelan performa akademik untuk mendukung pengambilan keputusan yang lebih objektif [3].

Akan tetapi, masih terdapat celah penelitian (*research gap*) yang signifikan, meskipun riset di bidang ini telah banyak dilakukan. Mayoritas penelitian sebelumnya cenderung hanya mengandalkan nilai akademik, seperti nilai rapor atau nilai ujian, sebagai fitur utama mengklasifikasikan performa siswa. Berdasarkan temuan Kotjek dkk. (2025), variabel non-kognitif seperti kebiasaan belajar dan gaya hidup justru memiliki korelasi yang kuat terhadap pencapaian nilai ujian siswa [4]. Faktor-faktor seperti jam belajar harian, durasi tidur,

tingkat kecemasan, hingga paparan *screen time* sering kali diabaikan dalam model klasifikasi, padahal faktor-faktor tersebut terbukti berpengaruh signifikan terhadap kondisi psikologis dan fokus belajar. Sejalan dengan pendapat Ahmed (2024), prediksi performa akademik yang akurat memerlukan pemahaman menyeluruh terhadap faktor-faktor yang memengaruhi siswa di luar lingkungan sekolah guna menghasilkan strategi evaluasi yang lebih proaktif [5].

Urgensi penelitian ini semakin nyata apabila ditinjau dari capaian pendidikan nasional terkini. Berdasarkan *Rapor Pendidikan Indonesia 2025* yang melaporkan hasil Asesmen Nasional (2024), proporsi murid yang mencapai kompetensi minimum literasi baru sebesar 70,03% dan numerasi 67,94%. Artinya, sekitar sepertiga siswa di Indonesia masih belum memenuhi kompetensi dasar [6]. Kondisi tersebut diperberat oleh tingginya beban pemantauan guru. Data *Statistik Pendidikan 2025* menunjukkan rasio murid-guru nasional berada pada kisaran 20:1 dari jenjang SD hingga SMA, sehingga deteksi dini kesulitan belajar pada tingkat individu sulit dilakukan secara manual [7]. Selain itu, pada tahun 2025 tercatat sebanyak 2.922.607 anak usia 7-18 tahun berstatus tidak bersekolah [8].

Sebagai solusi atas permasalahan tersebut, penelitian ini mengusulkan sebuah klasifikasi kemampuan siswa yang mengintegrasikan fitur akademik dan fitur non-kognitif menggunakan algoritma *K-Nearest Neighbor* (KNN). Berdasarkan pendapat dari Hidayatullah dkk. (2022), algoritma *K-Nearest Neighbor* (KNN) memiliki keunggulan pada kesederhanaan implementasi dan efektivitasnya dalam mengklasifikasikan data multidimensi tanpa memerlukan asumsi distribusi data yang kompleks [9]. Dalam penelitian ini, siswa akan diklasifikasi ke dalam tiga kategori utama, yaitu Lemah, Sedang, Pintar. Untuk memvalidasi bahwa algoritma KNN memberikan performa yang optimal pada konteks penelitian ini algoritma *Decision Tree* dan algoritma *Naive Bayes* digunakan sebagai baseline pembanding, perbandingan ini bertujuan untuk mengkonfirmasi keunggulan KNN secara empiris pada dataset yang digunakan, bukan untuk mencari algoritma terbaik secara terbuka. Output dari sistem ini tidak hanya berhenti pada kategori klasifikasi, melainkan menghasilkan rekomendasi pembelajaran yang bersifat personal dan disesuaikan dengan profil spesifik tiap kategori siswa.

Berdasarkan urgensi untuk menghasilkan pendekatan evaluasi yang lebih menyeluruh, penelitian ini penting dilakukan guna membantu sektor pendidikan dalam mendeteksi hambatan belajar secara dini melalui pola gaya hidup dan pencapaian akademik siswa. Oleh karena itu, penelitian ini mengimplementasikan

algoritma K-Nearest Neighbor untuk mengklasifikasikan kemampuan siswa sebagai dasar rekomendasi pembelajaran personal. Fokus penelitian ini diharapkan dapat memberikan kontribusi nyata dalam memperkuat sistem pendukung keputusan di bidang pendidikan melalui analisis data yang lebih personal dan komprehensif.

1.2 Rumusan Masalah

Berdasarkan Latar Belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana efektivitas algoritma *K-Nearest Neighbor* dalam mengklasifikasikan kemampuan siswa ke dalam kategori Lemah, Sedang, dan Pintar dengan mengintegrasikan fitur akademik dan fitur non-kognitif?
2. Bagaimana perbandingan performa model klasifikasi antara algoritma *K-Nearest Neighbor* dengan algoritma *Decision Tree* dan Algoritma *Naive Bayes* sebagai baseline dalam mengukur akurasi, presisi, dan *recall*?

1.3 Batasan Permasalahan

Guna menjaga fokus dan kedalaman analisis dalam penelitian ini, ditetapkan batasan-batasan masalah sebagai berikut:

1. Penelitian ini menggunakan dataset publik "*Student Performance Interactions*" dari platform Kaggle dengan tautan <https://www.kaggle.com/datasets/mabubaksiddiq/student-exam-performance/data>, yang mencakup total 1.000 data siswa sebagai basis simulasi.
2. Klasifikasi kemampuan siswa dibatasi pada tiga kategori utama, yaitu: Lemah (mencakup *grade* D dan F), Sedang (mencakup *grade* C), dan Pintar (mencakup *grade* A dan B). Penelitian tidak melakukan klasifikasi secara individual untuk setiap tingkatan *grade* (A-F) untuk menjaga fokus pada efektivitas pengelompokan tingkat kemampuan umum.
3. Algoritma klasifikasi yang diimplementasikan terbatas pada *K-Nearest Neighbor* (KNN) sebagai metode utama dan *Decision Tree* sebagai metode pembandingan. Penelitian tidak mencakup penggunaan metode *deep learning*, *ensemble learning*, atau algoritma optimasi lainnya.

4. Sistem rekomendasi pembelajaran adaptif yang dihasilkan bersifat *rule-based* (berbasis aturan) yang diderivasi langsung dari hasil analisis statistik deskriptif dataset. Batas ambang (*threshold*) rekomendasi ditentukan secara empiris berdasarkan rata-rata pola kelompok dan belum merujuk pada standar baku ilmu psikologi pendidikan maupun pedagogi nasional. Sistem ini juga belum mampu mendeteksi kelemahan spesifik per mata pelajaran secara individual.
5. Model tetap menyertakan fitur *parent_education_level* dan *study_environment* meskipun hasil analisis korelasi menunjukkan pengaruh yang sangat rendah terhadap performa siswa ($r = -0,03$ dan $r = -0,02$). Penelitian ini tidak melakukan tahap *feature selection* lanjutan untuk mengeliminasi fitur-fitur tersebut guna menguji performa algoritma terhadap struktur dataset asli secara utuh.
6. Penelitian ini berfokus pada pengembangan model klasifikasi dan simulasi logika rekomendasi, sehingga tidak mencakup implementasi ke dalam aplikasi berbasis web atau *mobile*, serta tidak melakukan integrasi ke dalam Sistem Informasi Akademik (SIKAD) yang sudah ada di institusi tertentu.

1.4 Tujuan Penelitian

Tujuan yang hendak dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengimplementasikan algoritma *K-Nearest Neighbor* untuk membangun model klasifikasi kemampuan siswa ke dalam kategori, Lemah, Sedang, dan Pintar. Sejalan dengan pendapat dari Hidayatullah dkk.(2022), implementasi ini ditujukan untuk menguji efektivitas KNN dalam menangani dan memproses fitur dari beberapa aspek yang mengintegrasikan data akademik dan non-kognitif [9].
2. Mengevaluasi dan memvalidasi apakah performa algoritma *K-Nearest Neighbor* terbukti menghasilkan kinerja yang lebih baik dari algoritma pembanding yaitu *Decision Tree* dan *Naive Bayes*. Mengacu pada metode penelitian yang membandingkan dua algoritma atau lebih yang di terapkan oleh Novianto dkk.(2023), serta Amri dkk.(2023), pengujian ini diukur secara percobaan nyata berdasarkan nilai metrik akurasi, *precision*, *recall*, dan *f1-score* untuk mengonfirmasi bahwa algoritma *K-Nearest Neighbor* pada

dataset yang digunakan memiliki ketangguhan dari algoritma pembandingan [10] [11].

3. Menghasilkan skema rekomendasi pembelajaran personal bagi setiap kategori tingkat kemampuan siswa. Berdasarkan temuan dari Pal dkk.(2025), dan Atalla dkk.(2023), perumusan aturan keputusan ini didasarkan pada pola rill hasil analisis statistik deksriptif dataset untuk memberikan rekomendasi tindakan yang relevan dengan profil spesifik dan hambatan belajar siswa [2] [3].
4. Mengidentifikasi dan menganalisis fitur akademik dan non-kognitif yang memberikan kontribusi paling signifikan terhadap akurasi model klasifikasi kemampuan siswa. Berdasarkan keterbatasan fitur yang diidentifikasi oleh Azizah dkk.(2022), langkah ini bertujuan untuk memetakan variabel mana saja (seperti *sleep hours*, *exam anxiety*, atau nilai mata pelajaran) yang memiliki pengaruh paling dominan dalam menentukan pengelompokan kemampuan belajar siswa [12].

1.5 Manfaat Penelitian

Manfaat yang diharapkan dapat diperoleh dari pelaksanaan penelitian ini adalah sebagai berikut:

1.5.1 Manfaat Teoretis (Akademis)

1. Menyediakan bukti nyata penggunaan algoritma *K-Nearest Neighbor* pada fitur non-kognitif dengan memberikan bukti nyata dan terukur mengenai bagaimana algoritma ini bekerja ketika diintegrasikan dengan fitur-fitur non-kognitif (seperti *sleep hours*, *exam anxiety*, atau waktu penggunaan gawai). Mengacu pada pendapat dari Hidayatullah dkk.(2022), algoritma *K-Nearest Neighbor* dikenal sangat efektif dalam memproses data yang memiliki banyak dimensi [9]. Melalui penelitian ini, para peneliti di bidang ilmu komputer dapat melihat secara langsung bagaimana performa algoritma *K-Nearest Neighbor* dalam mengolah pola perilaku sehari-hari siswa untuk menentukan kelompok kemampuan mereka secara tepat.
2. Menjadi referensi perbandingan algoritma *K-Nearest Neighbor* vs *Decision Tree* vs *Naive Bayes* dalam menghasilkan laporan pengujian yang jelas

mengenai perbandingan performa antara algoritma *K-Nearest Neighbor* sebagai metode utama melawan dua algoritma yang bertindak sebagai baseline, yaitu *Decision Tree* dan *Naive Bayes*. Sejalan dengan pendapat Novianto dkk.(2023) yang pernah menguji algoritma tersebut, hasil evaluasi dalam penelitian ini (berupa angka akurasi, *precision*, *recall*, dan *f1-score*) dapat menjadi referensi perbandingan yang valid bagi akademisi atau penelitian lain yang ingin melakukan studi komparasi sejenis di bidang data mining pendidikan.

3. Berkontribusi mengenai *Education Data Mining* dengan membongkar pengaruh variabel non-kognitif terhadap performa siswa. Hasil pengujian ini memperjelas sejauh mana faktor psikologis, gaya hidup, dan lingkungan berkontribusi dalam membentuk keputusan klasifikasi, sehingga menjadi tambahan wawasan baru bagi pengembangan model prediksi kecerdasan buatan di masa depan.

1.5.2 Manfaat Praktis

1. Sistem ini membantu para guru untuk mengelompokan tingkat kemampuan siswa dalam kategori (Lemah, Sedang, dan Pintar) secara otomatis, cepat, dan objektif. Guru tidak perlu lagi melakukan analisa manual yang memakan waktu atau sekedar mengira-ngira kemampuan siswa, karena model klasifikasi akan langsung memetakan data profil siswa secara terkomputerisasi berdasarkan gabungan data perilaku dan nilai akademis mereka.
2. Hasil klasifikasi model ini dapat dijadikan dasar utama bagi Guru, Wali Kelas, atau Guru BK untuk mengambil tindakan nyata di kelas. Berdasarkan temuan Ahmed (2024), strategi evaluasi yang proaktif sangat dibutuhkan oleh institusi pendidikan untuk mendeteksi masalah belajar sejak dini [5]. Dengan mengetahui kelompok siswa, sekolah bisa langsung merancang strategi mengajar yang pas misalnya memberikan bimbingan khusus untuk kelompok Lemah atau memberikan pengayaan berupa soal (HOTS) bagi kelompok Pintar.
3. Memberikan pemahaman mendalam bagi pihak sekolah bahwa pencapaian nilai siswa di kelas tidak berdiri sendiri. Sekolah dapat melihat secara nyata bahwa faktor diluar kelas, seperti durasi tidur dan kecemasan ujian yang

tinggi, memiliki andil yang besar dalam performa akademis siswa di sekolah. Hal ini membantu sekolah untuk menciptakan kebijakan atau lingkungan belajar yang lebih mendukung terhadap psikologis siswa.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun secara sistematis untuk memberikan gambaran menyeluruh mengenai alur penelitian yang dilakukan. Penjelasan singkat mengenai isi dari masing-masing bab adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Bab ini menguraikan latar belakang permasalahan yang mendasari penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan. Bab ini juga menjelaskan celah penelitian (*research gap*) yang menjadi dasar novelty penelitian terkait fitur akademik dan non-kognitif dalam klasifikasi kemampuan siswa.

- Bab 2 LANDASAN TEORI

Bab ini berisi tinjauan pustaka dan teori-teori yang mendasari penelitian, meliputi konsep klasifikasi, algoritma *K-Nearest Neighbor* (KNN), *Decision Tree*, *Naive Bayes*, teknik normalisasi *MinMaxScaler*, metrik evaluasi model (akurasi, *precision*, *recall*, *f1-score*, *confusion matrix*), serta konsep sistem rekomendasi pembelajaran personal. Selain itu, bab ini juga membuat kajian penelitian terdahulu yang relevan sebagai acuan dan pembandingan penelitian ini.

- Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan kerangka kerja penelitian yang digunakan, mencakup jenis penelitian, sumber dan karakteristik dataset yang digunakan, tahapan preprocessing data (encoding, normalisasi, pembagian data training dan testing), implementasi algoritma KNN sebagai metode utama beserta proses penentuan nilai K optimal, implementasi *Decision Tree* dan *Naive Bayes* sebagai baseline pembandingan, proses, evaluasi model, serta menguraikan perancangan skema sistem rekomendasi pembelajaran personal berbasis hasil analisis statistik deskriptif dataset.

- Bab 4 HASIL DAN DISKUSI

Bab ini menyajikan hasil dan analisis dari seluruh tahapan penelitian,

meliputi hasil eksplorasi data (EDA) dan analisis statistik deksriptif, hasil implementasi dan evaluasi model KNN dengan nilai K optimal, hasil perbandingan KNN dengan *Decision Tree* dan *Naive Bayes* berdasarkan metrik akurasi, *precision*, *recall*, dan *f1-score*, analisis fitur akademik dan non akademik yang memberikan kontribusi paling signifikan terhadap klasifikasi, serta skema rekomendasi pembelajaran personal yang dihasilkan untuk setiap kategori kemampuan siswa.

- Bab 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan akhir dari seluruh proses penelitian yang menjawab rumusan masalah serta memberikan saran-saran praktis bagi pihak terkait maupun rekomendasi bagi pengembangan penelitian selanjutnya.

