# BAB II

# KAJIAN PUSTAKA

## A. Tinjauan Pustaka

Bagian ini akan mengulas beberapa penelitian terdahulu yang menjadi fondasi ilmiah dalam pengembangan sistem rekomendasi spesialisasi mahasiswa. Penelaahan ini bertujuan untuk menunjukkan keterkaitan antara ide-ide peneliti sebelumnya dengan model integrasi dua tahap (K-Means dan CART) yang sedang dikembangkan dalam penelitian ini.

Pada penelitian (Sonianto & Hartono, 2025), Penelitian ini mengkaji permasalahan dalam menentukan minat dan bakat siswa sekolah dasar yang selama ini masih dilakukan secara manual dan cenderung subjektif. Kondisi tersebut dinilai kurang mendukung pembelajaran yang bersifat personal. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan siswa secara lebih objektif berdasarkan empat aspek utama, yaitu minat akademik, seni, olahraga, dan kepemimpinan. Metode yang digunakan adalah algoritma K-Means Clustering dengan validasi menggunakan Silhouette Score. Hasilnya menunjukkan bahwa siswa dapat dikelompokkan ke dalam tiga kelompok utama dengan karakter minat yang berbeda secara jelas. Namun demikian, penelitian ini masih memiliki keterbatasan karena hanya berhenti pada tahap pengelompokan, sehingga belum mampu memberikan rekomendasi otomatis yang dapat digunakan sebagai acuan langsung oleh siswa.

Penelitian oleh (Rose et al., 2024), Penelitian membahas permasalahan dalam proses pemilihan kelas di tingkat sekolah menengah yang masih dilakukan secara

konvensional dan memerlukan waktu yang cukup lama. Untuk mengatasi hal tersebut, penelitian ini mengembangkan aplikasi berbasis web dengan memanfaatkan algoritma K-Means Clustering yang diintegrasikan melalui Python (Tkinter). Hasil penelitian menunjukkan terbentuknya empat kelompok minat dengan nilai Silhouette Score sebesar 0,6233 yang menandakan kualitas pengelompokan yang baik. Sistem ini dinilai mampu membantu sekolah dalam memberikan rekomendasi kelas. Meskipun demikian, penelitian ini masih memiliki kekurangan karena belum menghasilkan aturan keputusan secara otomatis, sehingga hasil pengelompokan masih harus ditafsirkan secara manual oleh pihak sekolah.

Pada penelitian yang dilakukan oleh (Muhammad & Irvan, 2025), berfokus pada analisis minat belajar siswa berdasarkan mata pelajaran yang diminati dengan tujuan meningkatkan motivasi akademik. Penelitian ini menggunakan algoritma Decision Tree yang diimplementasikan dalam sistem berbasis web. Hasilnya menunjukkan bahwa sistem mampu menampilkan pola minat siswa secara jelas dan mudah dipahami. Namun, penelitian ini memiliki kelemahan pada keterbatasan variabel yang digunakan, karena hanya mengandalkan data akademik dasar dan belum mempertimbangkan aspek lain seperti aktivitas organisasi atau tingkat motivasi belajar yang lebih kompleks.

Penelitian yang dilakukan oleh (Betrand et al., 2025), membahas kesulitan mahasiswa dalam menentukan arah karier yang sesuai dengan minat dan kemampuan setelah lulus. Dalam penelitian ini digunakan beberapa algoritma klasifikasi seperti Decision Tree, Random Forest, dan Naïve Bayes untuk

menghasilkan rekomendasi karier. Hasil penelitian menunjukkan bahwa sistem rekomendasi dapat memberikan dampak positif bagi mahasiswa, baik secara ekonomi maupun psikologis. Akan tetapi, penelitian ini memiliki kekurangan karena langsung melakukan proses klasifikasi tanpa melalui tahap pengelompokan karakter terlebih dahulu, sehingga rekomendasi yang dihasilkan berpotensi kurang personal.

Penelitian oleh (Rahmallia, 2024), berupaya mengoptimalkan proses pemetaan potensi individu dengan mengolah data nilai akademik serta aktivitas ekstrakurikuler siswa menggunakan algoritma Random Forest dan Decision Tree. Penelitian ini menunjukkan bahwa penerapan metode machine learning mampu meningkatkan tingkat akurasi dalam mengidentifikasi minat dan bakat hingga mencapai 84%. Hal ini menandakan bahwa pendekatan berbasis data dapat memberikan hasil yang lebih objektif dibandingkan metode konvensional. Namun demikian, penelitian ini masih memiliki keterbatasan karena ruang lingkupnya hanya berada pada jenjang pendidikan menengah (SMA) yang karakteristik datanya relatif lebih sederhana. Variabel yang digunakan belum mencerminkan kompleksitas kebutuhan di tingkat perguruan tinggi

Pada penelitian yang dilakukan oleh (Sidik et al., 2025), mencoba mengembangkan sistem prediksi minat menggunakan tes kepribadian OCEAN dan tes aptitude dengan algoritma Random Forest yang diimplementasikan melalui aplikasi web berbasis Streamlit. Sistem ini mampu mencapai akurasi yang sangat tinggi, yaitu sebesar 97,42%. Meskipun demikian, penelitian ini lebih menekankan pada aspek kepribadian dan belum mengaitkannya dengan pengalaman nyata atau

performa akademik mahasiswa, sehingga hasil rekomendasi belum sepenuhnya mencerminkan kondisi real pengguna.

Penelitian oleh (Yağcı, 2022), membahas tentang efektivitas penggunaan algoritma machine learning untuk menemukan hubungan tersembunyi dalam data pendidikan guna memprediksi keberhasilan studi mahasiswa. Penelitian ini bertujuan untuk memprediksi nilai ujian akhir mahasiswa berdasarkan data nilai ujian tengah semester sebagai sumber data utama. Penelitian menggunakan metode perbandingan beberapa algoritma seperti *Random Forest*, *Naïve Bayes*, *SVM*, dan *KNN*. Hasil penelitian menunjukkan bahwa model yang diusulkan mencapai tingkat akurasi klasifikasi antara 70-75%. Kesimpulannya, pemanfaatan data pendidikan sangat penting untuk mendeteksi risiko kegagalan akademik lebih awal, namun riset ini belum diarahkan pada pemberian rekomendasi jalur spesialisasi karir yang spesifik berdasarkan pola minat mahasiswa.

Berdasarkan hasil telaah terhadap penelitian terdahulu, masih terdapat kesenjangan penelitian (*research gap*) dalam pengembangan sistem rekomendasi berbasis *machine learning*. Sebagian besar penelitian yang menerapkan K-Means Clustering hanya berfokus pada proses pengelompokan karakteristik pengguna tanpa menghasilkan rekomendasi yang dapat dimanfaatkan secara langsung. Sebaliknya, penelitian yang menggunakan Decision Tree umumnya melakukan proses klasifikasi secara langsung terhadap data berlabel sehingga belum mempertimbangkan karakteristik alami yang terbentuk dari data tanpa label. Akibatnya, rekomendasi yang dihasilkan belum sepenuhnya mampu

merepresentasikan pola perilaku maupun karakteristik pengguna secara lebih personal.

Berdasarkan kondisi tersebut, penelitian ini mengusulkan pendekatan hybrid machine learning yang mengintegrasikan algoritma K-Means Clustering dan Decision Tree CART dalam satu alur analisis. Algoritma K-Means dipilih sebagai tahap awal karena data penelitian berupa karakteristik perilaku belajar mahasiswa, seperti kehadiran, partisipasi, motivasi, aktivitas di luar perkuliahan, serta pengalaman di bidang teknologi informasi, yang belum memiliki label (*unlabeled data*). Melalui proses clustering, sistem mampu mengidentifikasi pola kemiripan antar mahasiswa secara objektif sehingga terbentuk kelompok karakter belajar yang dapat dijadikan representasi awal sebelum dilakukan proses klasifikasi. Pendekatan ini dinilai sesuai untuk menghasilkan profil mahasiswa berdasarkan struktur alami data.

Selanjutnya, Decision Tree CART digunakan untuk mengubah hasil pengelompokan tersebut menjadi rekomendasi yang lebih spesifik dan mudah dipahami. Pemilihan algoritma ini didasarkan pada kemampuannya menghasilkan aturan keputusan (*decision rules*) yang bersifat transparan, sehingga proses pembentukan rekomendasi dapat dijelaskan secara logis kepada pengguna. Berbeda dengan penelitian sebelumnya, label cluster hasil K-Means dimanfaatkan sebagai fitur tambahan pada proses klasifikasi, sehingga model tidak hanya memanfaatkan karakteristik individu mahasiswa, tetapi juga mempertimbangkan informasi kelompok yang telah terbentuk pada tahap clustering. Integrasi tersebut diharapkan mampu menghasilkan rekomendasi minat akademik yang lebih personal, memiliki

interpretasi yang lebih baik, serta tetap mempertahankan kemudahan dalam proses pengambilan keputusan.

## B. Landasan Teori

1. Minat Akademik

    a. Konsep dan Definisi Minat

    Secara teoretis, minat dapat dipahami sebagai kecenderungan hati yang tinggi terhadap suatu subjek atau aktivitas tertentu yang muncul secara sadar. Untuk mendapatkan pemahaman yang komprehensif, berikut adalah tinjauan definisi minat menurut beberapa ahli terkemuka:

    1. Djaali (2007: 121, dikutip dalam jurnal penelitian (Solehah et al., 2022)) mengartikan minat sebagai sebuah perasaan lebih suka serta rasa keterikatan pada suatu hal tanpa adanya unsur paksaan. Hal ini menegaskan bahwa minat bersifat intrinsik, muncul secara sukarela dari dorongan internal individu untuk terlibat dalam suatu objek.
    2. Slameto (2013: 57, dikutip dengan jurnal yang sama (Solehah et al., 2022)) menjelaskan bahwa minat adalah kondisi saat seseorang merasa tertarik pada suatu kegiatan, yang ditandai dengan kecenderungan untuk memberikan perhatian secara konsisten dan mengenang aktivitas tersebut dengan perasaan senang. Minat di sini dipandang sebagai faktor yang menetap dalam kepribadian.
    3. Walgito (1981: 38, (Solehah et al., 2022)) berpendapat bahwa minat adalah kondisi mental saat individu memberikan perhatian penuh pada

suatu objek yang disertai dengan keinginan kuat untuk mendalami, memahami, dan membuktikan kebenaran objek tersebut melalui aktivitas nyata.

4. Muhibbin Syah (2009: 136, (Solehah et al., 2022)) menitikberatkan pada aspek gairah, di mana minat dipandang sebagai kecenderungan dan antusiasme yang tinggi atau keinginan yang besar terhadap sesuatu yang dianggap memiliki nilai guna bagi individu tersebut.
5. Hilgard (sebagaimana dikutip dalam (Sartika & Siburian, 2025)) menyatakan bahwa minat merupakan kecenderungan yang menetap dalam diri seseorang untuk senantiasa memperhatikan dan menikmati suatu aktivitas atau konten pembelajaran secara berkelanjutan, yang membedakannya dari perhatian sesaat.

Berdasarkan sintesis dari berbagai pendapat tersebut, minat dapat didefinisikan sebagai keterikatan emosional dan kognitif yang membuat seseorang merasa senang, fokus, dan bersedia meluangkan energi untuk terlibat aktif dalam suatu bidang secara terus-menerus.

b. Konsep Minat Akademik

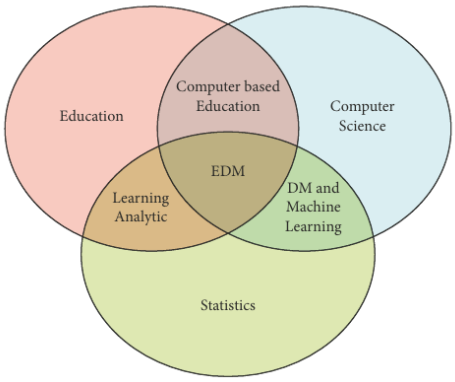
Berdasarkan sintesis dari berbagai pendapat mengenai minat, dapat disimpulkan bahwa minat merupakan keterikatan emosional dan kognitif yang mendorong individu untuk secara aktif terlibat dalam suatu aktivitas secara berkelanjutan. Dalam konteks pendidikan tinggi, konsep ini berkembang menjadi minat akademik, yaitu kecenderungan internal

mahasiswa untuk menunjukkan ketertarikan, perhatian, dan keterlibatan aktif terhadap bidang keilmuan tertentu dalam proses pembelajaran. Minat akademik tidak hanya mencerminkan rasa suka, tetapi juga mencakup dorongan untuk memahami, mendalami, serta mengembangkan kompetensi dalam bidang tersebut secara konsisten.

Sejalan dengan (Viona Sepahira et al., 2025), minat akademik berperan sebagai energi internal yang mengarahkan mahasiswa untuk berpartisipasi secara sadar dalam proses pendidikan guna mencapai tujuan profesional. Hal ini menunjukkan bahwa minat akademik tidak bersifat pasif, melainkan menjadi faktor pendorong utama dalam pembentukan arah spesialisasi keilmuan.

Dalam penelitian ini, minat akademik dipahami sebagai kecenderungan mahasiswa dalam memilih, menekuni, dan mengembangkan bidang keilmuan tertentu yang tercermin melalui perilaku belajar, preferensi mata kuliah, serta pengalaman akademik yang dimiliki. Dengan demikian, minat akademik tidak hanya dilihat sebagai konsep psikologis, tetapi juga sebagai variabel yang dapat diukur melalui indikator-indikator yang terstruktur.

2. Data Mining




Gambar 2. 1 Data Mining (Yağcı, 2022)

a. Konsep Dasar dan Filosofi Data Mining

Data Mining atau penambangan data secara teoretis dipahami sebagai cara sistematis untuk menggali informasi penting dari tumpukan data yang jumlahnya sangat melimpah atau masif. Merujuk pada pandangan (Fang, 2024), Data mining merupakan sebuah upaya sistematis untuk mendeteksi pola-pola yang bersifat implisit artinya pola tersebut ada namun tidak terlihat secara kasat mata serta harus teruji kebenarannya (valid) agar tidak hanya dianggap sebagai kebetulan belaka. Secara filosofis, bidang ilmu ini lahir sebagai jawaban atas fenomena "ledakan data", di mana dunia saat ini dibanjiri oleh informasi yang tumbuh sangat cepat namun sering kali "miskin" akan makna yang benar-benar bisa digunakan.

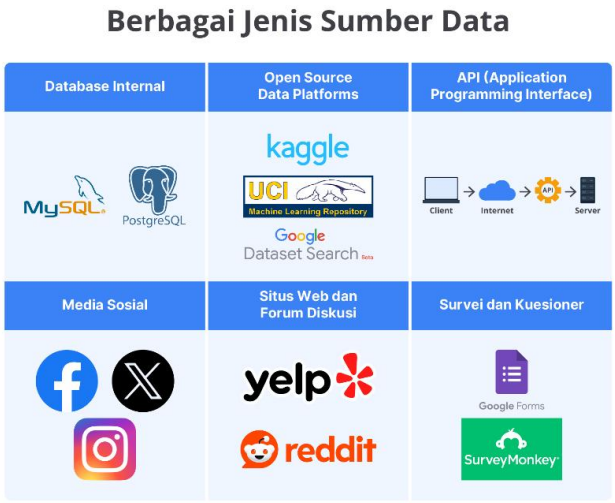
Dengan melibatkan bantuan algoritma pintar, sistem dapat mengenali "benang merah" atau struktur rahasia dalam riwayat masa lalu yang tidak mungkin bisa dikenali oleh nalar manusia secara manual. Hasil akhirnya

bukan sekadar statistik, melainkan sebuah peta pengetahuan yang kuat dan siap diimplementasikan dalam berbagai skenario dunia nyata guna memprediksi kecenderungan yang akan terjadi di masa depan.

b. Kerangka Kerja Knowledge Discovery in Database (KDD)

Proses pengolahan data mentah hingga menjadi suatu pengetahuan yang bermakna dilakukan melalui suatu alur kerja yang sistematis yang dikenal sebagai Knowledge Discovery in Database (KDD). Menurut (Fang, 2024), KDD merupakan proses penting dalam data mining yang terdiri dari beberapa tahapan utama yang saling berhubungan untuk memastikan bahwa hasil analisis yang diperoleh memiliki validitas dan dapat dipertanggungjawabkan secara ilmiah. Tahapan ini tidak hanya berfokus pada proses analisis data, tetapi juga mencakup proses persiapan data agar hasil yang diperoleh lebih akurat dan relevan.

1) Data Selection




*Gambar 2. 2 Berbagai jenis sumber data (Angel Metanosa Afinda, 2024)*

Tahap awal dalam proses KDD yang dilakukan dengan cara memilih dan menyaring data yang relevan dari keseluruhan basis data yang tersedia. Tujuan dari tahap ini adalah untuk menentukan subset data yang sesuai dengan kebutuhan analisis, sehingga proses pengolahan data dapat lebih terarah. Dengan adanya seleksi data, sistem hanya akan memproses variabel-variabel yang memiliki hubungan atau kontribusi terhadap tujuan penelitian, sehingga dapat mengurangi kompleksitas data dan meningkatkan efisiensi proses analisis (Fang, 2024). Tahap ini juga membantu menghindari masuknya data yang tidak relevan yang dapat memengaruhi hasil akhir.

2) Data Cleaning

Tahap yang berfungsi untuk meningkatkan kualitas data yang telah dikumpulkan sebelum dilakukan proses analisis lebih lanjut. Pada tahap ini, data diperiksa dan dibersihkan dari berbagai permasalahan seperti missing

values, data yang tidak konsisten, serta adanya outlier yang dapat mengganggu proses analisis. Menurut (Jo, 2021), kualitas data memiliki peran yang sangat penting dalam menentukan kualitas hasil akhir, karena data yang tidak valid dapat menghasilkan kesimpulan yang bias atau tidak akurat. Oleh karena itu, dilakukan proses seperti pengisian nilai yang hilang (imputasi), perbaikan data yang tidak sesuai, atau penghapusan data yang dianggap tidak relevan agar data menjadi lebih bersih dan siap untuk diproses (Fang, 2024).

3) Data Transformation

Tahap lanjutan setelah data dinyatakan bersih, di mana data diubah ke dalam bentuk yang sesuai dengan kebutuhan algoritma yang digunakan. Menurut (Jo, 2021) proses ini mencakup pengubahan data kategorikal atau teks menjadi data numerik melalui teknik encoding, karena sebagian besar algoritma machine learning hanya dapat bekerja dengan data berbentuk angka. Selain itu, dilakukan juga proses normalisasi atau penskalaan data seperti Min-Max Scaling agar setiap variabel memiliki rentang nilai yang seimbang. Hal ini penting untuk mencegah terjadinya dominasi variabel tertentu yang memiliki skala lebih besar terhadap hasil perhitungan model (Fang, 2024)

4) Data Mining/Modeling

Inti dari seluruh proses KDD, yaitu tahap di mana pengetahuan diekstraksi dari data menggunakan algoritma tertentu. Pada tahap ini, dilakukan proses pembelajaran model (training) menggunakan data yang telah diproses

sebelumnya untuk menemukan pola-pola tersembunyi yang terdapat dalam dataset. Dalam penelitian ini, digunakan algoritma K-Means untuk melakukan pengelompokan data berdasarkan kemiripan karakteristik, serta algoritma CART untuk membentuk aturan keputusan dalam proses klasifikasi. Tahap ini bertujuan untuk menemukan hubungan, pola, maupun struktur yang dapat merepresentasikan karakteristik data secara lebih bermakna (Fang, 2024).

5) Evaluation

Tahap akhir dalam proses KDD yang digunakan untuk menguji dan memvalidasi hasil model yang telah dibangun. Pada tahap ini, model diuji menggunakan data pengujian (testing data) untuk menilai sejauh mana model mampu menghasilkan prediksi yang sesuai dengan kondisi sebenarnya. Evaluasi dilakukan dengan menggunakan beberapa metrik seperti akurasi, presisi, serta tingkat error (error rate) untuk menentukan tingkat keberhasilan model. Hasil evaluasi ini menjadi dasar dalam menentukan apakah model sudah layak digunakan atau masih memerlukan perbaikan lebih lanjut pada parameter tertentu (Fang, 2024).

## 3. Machine Learning

a. Konsep Machine Learning

Machine Learning (ML) merupakan salah satu cabang *Artificial Intelligence* yang memungkinkan komputer membangun model melalui proses pembelajaran dari sekumpulan data sehingga mampu mengenali

pola, melakukan prediksi, maupun menghasilkan keputusan terhadap data baru. Menurut Jo (2021), konsep utama Machine Learning terletak pada kemampuan model untuk memperoleh pengetahuan dari contoh-contoh yang tersedia (*training examples*), kemudian melakukan generalisasi terhadap data yang belum pernah dipelajari sebelumnya. Dengan demikian, sistem tidak lagi bergantung sepenuhnya pada aturan yang ditentukan secara eksplisit oleh manusia, tetapi membentuk pola keputusan berdasarkan informasi yang dipelajari dari data.

Secara umum, Machine Learning mencakup beberapa paradigma pembelajaran, antara lain supervised learning, unsupervised learning, semi-supervised learning, dan reinforcement learning. Pada *supervised learning*, model mempelajari hubungan antara data masukan dan target yang telah memiliki label. Sebaliknya, *unsupervised learning* digunakan ketika data belum memiliki label sehingga model bertugas menemukan struktur atau pola alami berdasarkan tingkat kemiripan antar data.

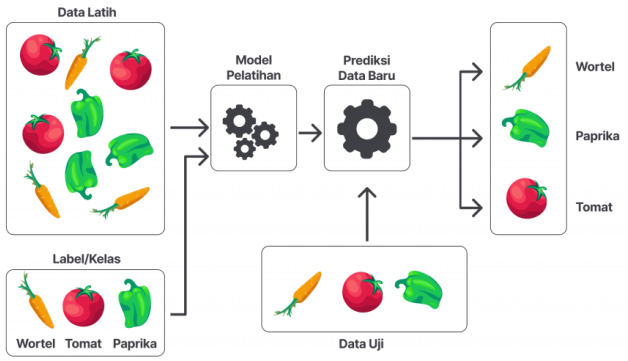
Jumlah data bukan merupakan faktor tunggal yang menentukan penerapan Machine Learning. Kinerja model dipengaruhi oleh berbagai aspek, seperti kualitas data, representativitas sampel, kompleksitas permasalahan, jumlah fitur, serta algoritma yang digunakan. Oleh karena itu, kebutuhan jumlah data dapat berbeda pada setiap kasus. Pada permasalahan yang relatif sederhana, model dapat dibangun menggunakan dataset berukuran kecil apabila data tersebut mampu merepresentasikan

karakteristik permasalahan secara memadai, sedangkan pada permasalahan yang lebih kompleks diperlukan jumlah data yang lebih besar (Jo, 2021).

b. Jenis Pembelajaran *Machine Learning*

Dalam bidang kecerdasan buatan dan *machine learning*, proses analisis data dapat dilakukan melalui beberapa pendekatan pembelajaran. Berdasarkan literatur dari (Jo, 2021) dan (Fang, 2024), metode pembelajaran dalam sistem cerdas secara umum dapat dibedakan menjadi dua paradigma utama, yaitu Supervised Learning dan Unsupervised Learning. Setiap paradigma memiliki strategi nalar yang spesifik untuk mengubah data mentah menjadi sebuah kesimpulan yang bermakna:

1) *Supervised Learning*




Gambar 2. 3 Supervised Learning *(Angel Metanosa Afinda, 2024a)*

*Supervised learning* merupakan pendekatan dalam machine learning yang memanfaatkan data yang telah memiliki label atau target sebagai dasar proses pembelajaran model. Dalam pendekatan ini, setiap data yang digunakan sudah disertai dengan jawaban yang benar sehingga sistem dapat mempelajari hubungan antara karakteristik data masukan (*input*) dengan hasil yang diharapkan (*output*).Melalui proses pelatihan yang dilakukan secara berulang, sistem secara bertahap mampu mengenali pola yang menghubungkan atribut data dengan label yang sesuai. Tujuan utama dari proses ini adalah membentuk sebuah fungsi pemetaan yang mampu melakukan generalisasi pengetahuan sehingga model dapat memberikan prediksi yang tepat ketika dihadapkan pada data baru yang belum pernah dipelajari sebelumnya. (Jo, 2021)

Berdasarkan jenis tugas yang dilakukan, pembelajaran terbimbing umumnya dibagi menjadi dua kategori utama, yaitu klasifikasi (*classification*) dan regresi (*regression*).

A. Klasifikasi (*Classification*)

Klasifikasi merupakan metode yang digunakan ketika sistem bertugas menentukan kategori atau label dari suatu objek berdasarkan karakteristik yang dimilikinya. Hasil dari proses ini bersifat diskrit, artinya setiap data akan ditempatkan ke dalam salah satu kategori yang telah ditentukan sebelumnya.

1. Classification and Regression Tree (CART)

Menurut (Abdul Majid et al., 2022), algoritma Decision Tree CART (Classification and Regression Tree) merupakan salah satu metode klasifikasi dalam *data mining* yang digunakan untuk membentuk model keputusan dalam struktur pohon (*decision tree*). Model tersebut digunakan untuk memetakan hubungan antaratribut sehingga mampu menghasilkan aturan keputusan yang dapat digunakan sebagai dasar proses klasifikasi. Struktur Decision Tree terdiri atas tiga komponen utama, yaitu root node, branch, dan leaf node. *Root node* merupakan simpul awal yang berisi atribut utama sebagai dasar pemisahan data, *branch* menunjukkan hasil percabangan berdasarkan atribut yang diuji, sedangkan *leaf node* merupakan simpul akhir yang menyajikan hasil klasifikasi atau keputusan dari data yang dianalisis (Abdul Majid et al., 2022).

Berbeda dengan beberapa algoritma *decision tree* lainnya, CART membangun pohon keputusan menggunakan mekanisme binary split, yaitu setiap simpul hanya menghasilkan dua cabang. Pendekatan tersebut menghasilkan struktur pohon yang lebih sederhana sehingga proses pengambilan keputusan menjadi lebih mudah dipahami dan diinterpretasikan. Selain itu, CART umumnya menggunakan Gini Index sebagai ukuran impurity untuk menentukan atribut terbaik pada setiap percabangan sehingga mampu menghasilkan model klasifikasi yang efisien dan memiliki kemampuan prediksi yang baik (Abdul Majid et al., 2022).

Dalam penerapannya, algoritma Decision Tree CART telah banyak dimanfaatkan pada berbagai bidang penelitian, termasuk bidang pendidikan. Metode ini digunakan untuk memprediksi berbagai permasalahan akademik, seperti penentuan minat studi lanjutan, klasifikasi karakteristik mahasiswa, maupun analisis faktor yang memengaruhi prestasi belajar. Penelitian yang dilakukan oleh (Irmayanti et al., 2025) menunjukkan bahwa algoritma CART mampu menghasilkan tingkat akurasi yang baik sehingga layak digunakan sebagai metode pendukung pengambilan keputusan berbasis data.

Pada penelitian ini, pembentukan pohon keputusan dilakukan menggunakan ukuran Entropy sebagai kriteria pemilihan atribut terbaik pada setiap percabangan, sehingga proses klasifikasi diawali dengan menghitung nilai Entropy dan Information Gain untuk menentukan atribut yang paling efektif dalam memisahkan data. Mekanisme tersebut bertujuan menghasilkan struktur pohon keputusan yang mampu mengelompokkan data secara optimal sehingga proses klasifikasi menjadi lebih akurat dan mudah diinterpretasikan (Fang, 2024).

Berdasarkan landasan teori yang dikemukakan oleh (Fang, 2024) serta didukung oleh penelitian (Irmayanti et al., 2025), mekanisme kerja algoritma ini dibangun atas tiga komponen utama, yaitu:

a) Entropy (Indikator Kekacauan/Impurity)

Entropy merupakan ukuran yang digunakan untuk mengetahui tingkat keberagaman atau ketidakseragaman data dalam suatu kelompok. Nilai entropy yang tinggi menunjukkan bahwa data dalam kelompok tersebut masih beragam atau belum memiliki pola yang jelas. Sebaliknya, nilai entropy yang rendah menunjukkan bahwa data dalam kelompok cenderung seragam dan memiliki kecenderungan yang sama. Dalam penelitian ini, entropy digunakan sebagai dasar untuk menilai apakah data sudah cukup jelas untuk dijadikan acuan dalam proses pengambilan keputusan atau masih memerlukan pemisahan lebih lanjut.

Rumus Entropy:

$$\text{Entropy(S)} = \sum_{\text{i=1}}^{\text{n}} -\text{p}_\text{i} \log_2 \text{p}_\text{i}$$

(Abdul Majid et al., 2022)

Keterangan:

1. $(S)$ merupakan kumpulan data yang sedang dianalisis
2. $p_i$ adalah proporsi anggota kelompok yang memiliki minat tertentu.

Sebagai contoh, dalam satu kelompok mahasiswa terdapat perbedaan karakteristik yang cukup beragam, seperti tingkat keaktifan, pengalaman proyek, dan gaya belajar yang tidak

seragam. Kondisi ini menunjukkan bahwa data masih bercampur sehingga nilai entropy cenderung tinggi. Sebaliknya, jika dalam satu kelompok mayoritas mahasiswa memiliki karakteristik yang mirip, seperti sama-sama aktif dan memiliki pengalaman praktik, maka data menjadi lebih terarah dan nilai entropy akan lebih rendah.

Dalam penelitian ini, nilai entropy digunakan sebagai titik awal untuk melihat tingkat keragaman data mahasiswa berdasarkan jawaban yang diberikan pada kuesioner. Dengan mengetahui tingkat keragaman tersebut, sistem dapat menentukan langkah selanjutnya dalam proses pembentukan pohon keputusan untuk memberikan rekomendasi yang lebih tepat.

b) Information Gain (Menentukan Pertanyaan Paling Berpengaruh)

Information Gain merupakan ukuran yang digunakan untuk mengetahui seberapa besar suatu atribut mampu mengurangi tingkat ketidakpastian dalam data. Setelah nilai entropy dari seluruh data diketahui, algoritma CART akan menguji setiap atribut yang tersedia untuk melihat atribut mana yang paling efektif dalam memisahkan data menjadi kelompok yang lebih jelas.

Rumus Information Gain:

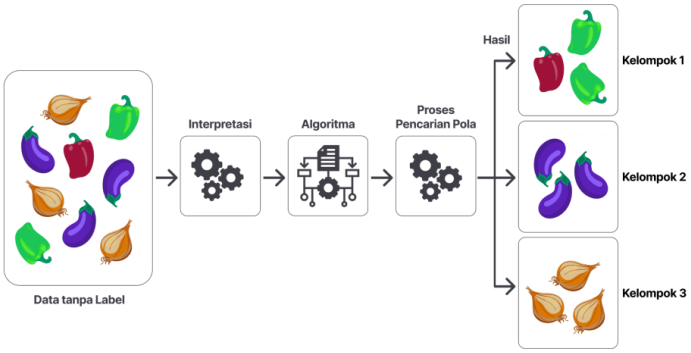
$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

(Abdul Majid et al., 2022)

Keterangan:

1. $S$ adalah keseluruhan data
2. $A$ adalah atribut yang diuji
3. $S_v$adalah subset data hasil pembagian berdasarkan atribut $A$
4. $|S_v|$ adalah jumlah data pada subset ke
5. $|S|$adalah jumlah seluruh data.

2) Unsupervised Learning




Gambar 2. 4 Unsupervised Learning *(Angel Metanosa Afinda, 2024a)*

*Unsupervised learning* merupakan pendekatan dalam *machine learning* yang memungkinkan sistem mempelajari pola dari data tanpa adanya label atau target jawaban yang telah ditentukan sebelumnya. Dalam pendekatan ini, sistem diberikan kumpulan data mentah dan diminta untuk menemukan sendiri pola, keteraturan, atau kemiripan yang terdapat di dalam data tersebut. Berbeda dengan *supervised learning*, metode ini tidak memiliki acuan hasil yang benar, sehingga sistem melakukan proses eksplorasi secara mandiri berdasarkan karakteristik data yang tersedia. Secara umum, menurut (Jo, 2021) dan (Fang, 2024) pendekatan *unsupervised learning* dapat dibagi ke dalam beberapa kategori, di antaranya klasterisasi (*clustering*) dan asosiasi (*association*).

A. Klasterisasi (Clustering)

Klasterisasi merupakan teknik yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik

yang dimiliki. Data yang memiliki karakteristik serupa akan ditempatkan dalam satu kelompok yang sama, sedangkan data yang berbeda akan berada pada kelompok yang berbeda. Tujuan utama dari proses klasterisasi adalah menemukan struktur alami dalam data tanpa adanya informasi kategori sebelumnya.

1. K-Means Clustering

Algoritma K-Means Clustering merupakan salah satu metode *unsupervised learning* dalam *Machine Learning* yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik antarobjek. Tujuan utama algoritma ini adalah membagi sekumpulan data ke dalam beberapa kelompok (*cluster*) sehingga data yang berada pada *cluster* yang sama memiliki tingkat kemiripan yang tinggi, sedangkan data pada *cluster* yang berbeda mempunyai karakteristik yang relatif berbeda (Sitti Rahmah et al., 2025). Metode ini sesuai digunakan pada data yang belum memiliki label karena mampu menemukan pola alami (*natural grouping*) tanpa memerlukan informasi kelas sebelumnya.

K-Means termasuk algoritma *centroid-based clustering*, yaitu setiap *cluster* direpresentasikan oleh sebuah titik pusat (*centroid*). Pendekatan ini dipilih karena memiliki mekanisme yang relatif sederhana, efisien secara komputasi, serta mudah diinterpretasikan, khususnya pada data numerik. Selain itu,

algoritma ini berupaya meminimalkan jarak antara setiap data dengan *centroid* kelompoknya sehingga menghasilkan kelompok yang memiliki tingkat homogenitas yang lebih baik(Abdul Majid et al., 2022).

Proses kerja algoritma K-Means diawali dengan menentukan jumlah *cluster* (K) yang akan dibentuk. Selanjutnya sistem menetapkan *centroid* awal sebagai pusat masing-masing *cluster*. Setiap data kemudian dihitung jaraknya terhadap seluruh *centroid* menggunakan metode Euclidean Distance, kemudian ditempatkan pada *cluster* dengan jarak terdekat. Setelah seluruh data memperoleh kelompok, posisi *centroid* diperbarui berdasarkan nilai rata-rata anggota pada setiap *cluster*. Tahapan tersebut dilakukan secara berulang (*iterative process*) hingga posisi *centroid* tidak lagi mengalami perubahan atau telah mencapai kondisi konvergen sehingga diperoleh hasil pengelompokan yang optimal (Abdul Majid et al., 2022)

Dalam penelitian ini, algoritma K-Means dipilih karena data karakteristik mahasiswa yang digunakan belum memiliki label (*unlabeled data*), sehingga diperlukan metode yang mampu mengidentifikasi pola kemiripan secara otomatis. Hasil pengelompokan tersebut selanjutnya dimanfaatkan sebagai dasar pembentukan *pseudo-label* melalui pendekatan Weak Supervision, sebelum digunakan pada proses klasifikasi

menggunakan algoritma Decision Tree CART. Dengan demikian, proses pengelompokan tidak hanya berfungsi untuk menemukan karakteristik mahasiswa, tetapi juga mendukung pembentukan model rekomendasi yang lebih sistematis dan mudah diinterpretasikan.

1) Perhitungan Jarak Menggunakan Euclidean Distance

Dalam proses pengelompokan data menggunakan algoritma K-Means, sistem perlu mengetahui tingkat kemiripan antar data. Untuk menentukan kemiripan tersebut, diperlukan suatu metode yang dapat mengukur jarak antara satu data dengan data lainnya. Salah satu metode yang paling umum digunakan adalah Euclidean Distance.

Secara konsep, Euclidean Distance merupakan metode yang digunakan untuk menghitung jarak antara dua titik dalam suatu ruang data. Prinsipnya sama seperti menghitung jarak antara dua titik pada bidang koordinat. Semakin kecil jarak yang dihasilkan, maka semakin tinggi tingkat kemiripan antara kedua data tersebut. Sebaliknya, semakin besar jaraknya, maka semakin berbeda karakteristik data tersebut.

Perhitungan jarak Euclidean dinyatakan dengan rumus sebagai berikut:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - c_{jk})^2}$$

Keterangan:

1. $d(x_i, c_j)$ : Jarak antara data ke i dengan centroid kelompok ke j
2. $x_i$ : Data yang akan dihitung jaraknya
3. $c_j$ : Pusat kelompok (*centroid*)
4. $n$ : Jumlah variabel dalam data

Secara sederhana, rumus ini bekerja dengan menghitung selisih nilai setiap variabel antara data dengan pusat kelompok, kemudian menjumlahkan seluruh selisih tersebut untuk memperoleh jarak total.

2) Titik Keseimbangan (Centroid Update)

Setelah seluruh data ditempatkan ke dalam kelompok berdasarkan jarak terdekat, langkah berikutnya dalam algoritma K-Means adalah memperbarui posisi pusat kelompok atau yang disebut centroid.

Centroid merupakan titik yang mewakili rata-rata karakteristik dari seluruh anggota dalam suatu kelompok. Oleh karena itu, setelah proses pengelompokan awal

dilakukan, posisi centroid perlu dihitung kembali agar dapat mencerminkan kondisi data yang sebenarnya.

Perhitungan centroid dilakukan dengan menggunakan rumus rata-rata sebagai berikut:

$$v_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_i$$

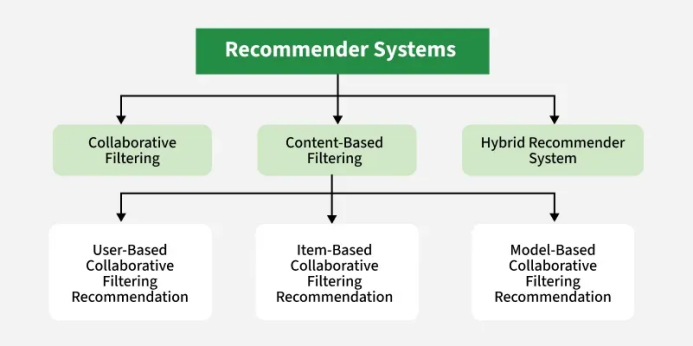
Keterangan:

1. $v_j$ : Posisi centroid baru pada kelompok ke-j
2. $N_j$ : Jumlah data dalam kelompok ke-j
3. $x_i$ : Data anggota kelompok

Secara sederhana, rumus ini menghitung nilai rata-rata dari seluruh data yang berada dalam kelompok tersebut. Nilai rata-rata tersebut kemudian digunakan sebagai posisi pusat kelompok yang baru.

Sebagai contoh, apabila suatu kelompok terdiri dari beberapa mahasiswa yang memiliki tingkat kehadiran dan jam belajar yang relatif tinggi, maka centroid kelompok tersebut juga akan berada pada nilai yang tinggi. Dengan demikian, centroid dapat mewakili karakteristik umum dari anggota kelompok tersebut.

c. Sistem Rekomendasi (Recommendation System)




Gambar 2. 5 Sistem Rekomendasi (GeeksforGeeks, 2025)

Dalam era informasi yang berkembang sangat pesat, manusia dihadapkan pada jumlah informasi yang sangat besar. Kondisi ini sering kali membuat individu mengalami kesulitan dalam menentukan pilihan yang paling sesuai dengan preferensi pribadinya. Dalam konteks tersebut, sistem rekomendasi hadir sebagai salah satu teknologi yang berfungsi membantu pengguna dalam menyaring dan memilih informasi yang relevan.

Menurut (Jo, 2021), sistem rekomendasi merupakan suatu sistem yang dirancang untuk memprediksi tingkat ketertarikan atau preferensi seorang pengguna terhadap suatu objek tertentu. Objek yang dimaksud dapat berupa berbagai hal, seperti produk, layanan, informasi, maupun konten digital. Secara konseptual, sistem rekomendasi bekerja dengan memanfaatkan pola perilaku pengguna pada masa lalu untuk memperkirakan kemungkinan preferensi di masa mendatang. Sistem ini mempelajari berbagai bentuk interaksi pengguna, baik yang bersifat eksplisit maupun implisit. Interaksi eksplisit dapat berupa pemberian nilai

atau rating, sedangkan interaksi implisit dapat berupa riwayat pilihan, aktivitas penggunaan, maupun pola perilaku lainnya.

1) Klasifikasi Jenis-Jenis Sistem Rekomendasi

Berdasarkan literatur yang dikemukakan oleh (Fang, 2024) dan (Jo, 2021), terdapat beberapa pendekatan yang umum digunakan dalam membangun sistem rekomendasi. Pendekatan-pendekatan tersebut digunakan untuk menghasilkan saran yang relevan sesuai dengan preferensi pengguna. Beberapa pendekatan utama dalam sistem rekomendasi antara lain sebagai berikut.

Berdasarkan literatur yang dikemukakan oleh (Fang, 2024) dan (Jo, 2021), terdapat beberapa pendekatan yang umum digunakan dalam membangun sistem rekomendasi. Pendekatan-pendekatan tersebut digunakan untuk menghasilkan saran yang relevan sesuai dengan preferensi pengguna. Beberapa pendekatan utama dalam sistem rekomendasi antara lain sebagai berikut.

*A. Content-Based Filtering*

Content-Based Filtering merupakan pendekatan yang bekerja dengan menganalisis karakteristik atau fitur dari item yang sebelumnya disukai oleh pengguna. Sistem akan mempelajari atribut dari item tersebut, kemudian mencari item lain yang memiliki karakteristik serupa untuk direkomendasikan kepada pengguna.

Sebagai contoh, apabila seorang pengguna sering memilih konten dengan tema teknologi, maka sistem akan merekomendasikan konten lain yang memiliki atribut atau topik yang serupa. Keunggulan pendekatan ini adalah kemampuannya dalam memberikan rekomendasi yang bersifat personal tanpa bergantung pada data pengguna lain. Namun, metode ini memiliki keterbatasan dalam menyajikan variasi rekomendasi yang benar-benar baru di luar pola minat pengguna sebelumnya.

*B. Collaborative Filtering*

Collaborative Filtering merupakan pendekatan yang memanfaatkan kemiripan pola perilaku antar pengguna. Metode ini bekerja dengan membandingkan riwayat interaksi atau preferensi antar individu untuk menemukan pengguna yang memiliki pola minat yang serupa. Sebagai contoh, apabila Pengguna A dan Pengguna B memiliki riwayat pilihan yang hampir sama, maka item yang disukai oleh Pengguna B namun belum pernah dipilih oleh Pengguna A dapat direkomendasikan kepada Pengguna A. Pendekatan ini cukup efektif dalam menemukan rekomendasi baru yang mungkin belum pernah dipertimbangkan oleh pengguna sebelumnya.

*C. Hybrid Recommendation System*

Sistem rekomendasi hybrid merupakan pendekatan yang menggabungkan dua atau lebih metode rekomendasi, biasanya antara Content-Based Filtering dan Collaborative Filtering. Penggabungan metode ini bertujuan untuk mengurangi kelemahan yang terdapat pada masing-

masing pendekatan jika digunakan secara terpisah. Dengan memanfaatkan kelebihan dari kedua metode tersebut, sistem rekomendasi hibrida mampu menghasilkan rekomendasi yang lebih akurat, stabil, dan relevan bagi pengguna.

*D. Knowledge-Based Recommendation*

Knowledge-Based Recommendation merupakan pendekatan yang tidak bergantung pada riwayat interaksi pengguna pada masa lalu. Metode ini menggunakan pengetahuan atau aturan tertentu yang telah ditentukan sebelumnya untuk memberikan rekomendasi. Rekomendasi diberikan berdasarkan kebutuhan pengguna yang dinyatakan secara langsung serta pengetahuan yang dimiliki oleh sistem mengenai hubungan antara kebutuhan tersebut dengan item yang tersedia. Pendekatan ini umumnya digunakan pada bidang yang memerlukan pertimbangan logis yang lebih kompleks.

E. Sistem Rekomendasi Sadar Konteks (Context-Aware Recommendation)

Context-Aware Recommendation merupakan pendekatan yang mempertimbangkan faktor konteks dalam proses pemberian rekomendasi. Faktor konteks dapat berupa waktu, lokasi, perangkat yang digunakan, maupun kondisi tertentu yang memengaruhi preferensi pengguna. Dengan mempertimbangkan informasi kontekstual tersebut,

sistem dapat memberikan rekomendasi yang lebih dinamis dan relevan sesuai dengan situasi yang sedang dialami oleh pengguna.

Dengan mempertimbangkan informasi kontekstual tersebut, sistem dapat memberikan rekomendasi yang lebih dinamis dan relevan sesuai dengan situasi yang sedang dialami oleh pengguna.

4. Weak Supervision (Heuristic Labeling)

Weak Supervision merupakan pendekatan dalam pembelajaran mesin yang dirancang untuk mengatasi keterbatasan ketersediaan data berlabel dengan cara menghasilkan label secara otomatis dari berbagai sumber supervisi yang bersifat lemah. Berdasarkan (Zhang et al., 2022), pendekatan ini mengombinasikan beberapa sumber informasi seperti aturan heuristik, model prediktif, maupun pengetahuan eksternal yang memiliki tingkat akurasi tidak sempurna sehingga dapat membentuk dataset pelatihan tanpa ketergantungan pada anotasi manual .

Dalam kerangka Programmatic Weak Supervision, pelabelan dilakukan melalui mekanisme *labeling functions*, yaitu fungsi berbasis aturan yang dirancang untuk memberikan label secara otomatis pada data tertentu. Setiap labeling function dapat menghasilkan label yang berbeda satu sama lain, sehingga berpotensi menimbulkan ketidakkonsistenan dalam hasil pelabelan.

Untuk mengatasi hal tersebut, seluruh keluaran labeling functions digabungkan menggunakan model agregasi atau label model yang menghasilkan label probabilistik yang lebih stabil. Label tersebut kemudian digunakan sebagai data pelatihan untuk model machine learning, sehingga

memungkinkan proses pembelajaran dilakukan meskipun data awal tidak memiliki anotasi manual.

5. Bootstrap

Bootstrap merupakan kerangka kerja (*framework*) berbasis sumber terbuka (*open source*) yang paling populer untuk mempermudah proses pengembangan antarmuka atau *front-end* sebuah website melalui integrasi HTML, CSS, dan JavaScript. Menurut (Miftahul Huda, 2020), Bootstrap adalah sebuah alat bantu yang menyediakan sekumpulan komponen desain siap pakai untuk meningkatkan efisiensi waktu pengerjaan proyek digital dengan menyediakan pola desain yang terstandarisasi.

Selain itu, penggunaan Bootstrap dalam sistem informasi akademik, sebagaimana dijelaskan oleh (Miftahul Huda, 2020), memiliki peran penting dalam memastikan tampilan website dapat menyesuaikan diri dengan berbagai ukuran perangkat. Dengan demikian, pengguna tetap dapat mengakses sistem dengan nyaman baik melalui komputer, tablet, maupun ponsel. Hal ini juga mendukung proses pengelolaan data yang lebih fleksibel dan efisien karena sistem dapat diakses kapan saja dan dari perangkat apa pun.

Cara kerja Bootstrap pada dasarnya dibangun dari beberapa konsep utama yang bertujuan untuk meningkatkan kenyamanan pengguna saat mengakses website (*user experience*). Konsep pertama adalah *Responsive Web Design* dengan pendekatan *Mobile-First*. Menurut (Miftahul Huda, 2020), pendekatan ini memprioritaskan tampilan website agar terlebih dahulu nyaman digunakan pada

perangkat seluler, kemudian menyesuaikan ke layar yang lebih besar seperti tablet atau komputer.

6. Flask

Flask adalah sebuah kerangka kerja (*web framework*) yang dibangun dengan bahasa pemrograman Python dan termasuk dalam kategori *micro-framework*. Menurut (Grinberg, 2018), istilah *micro-framework* pada Flask menunjukkan filosofi desain yang menjaga inti sistem tetap sederhana, tetapi tetap memungkinkan pengembang untuk menambahkan fitur tambahan sesuai kebutuhan.

Berbeda dengan framework berskala besar yang biasanya memiliki struktur dan komponen yang kaku, Flask tidak memaksa pengembang mengikuti pola tertentu atau menggunakan pustaka tambahan seperti basis data atau alat validasi tertentu. Hal ini memberikan kebebasan penuh bagi pengembang untuk merancang arsitektur sistem sesuai kebutuhan spesifik, sehingga sangat cocok digunakan dalam pengembangan aplikasi penelitian atau sistem informasi yang memerlukan fleksibilitas tinggi (Grinberg, 2018).

Secara operasional, Flask bekerja dengan memanfaatkan beberapa konsep utama yang saling terhubung sehingga sistem dapat berjalan secara dinamis. Menurut Miguel (Grinberg, 2018) dan (Ma'arif & Kurniasih, 2024), terdapat empat pilar penting yang mendukung cara kerja Flask, yaitu sebagai berikut:

1) Routing (Pengatur Jalur dan Navigasi)

Routing merupakan mekanisme yang menghubungkan alamat URL dengan

fungsi Python tertentu. Dengan menggunakan *@app.route()*, setiap permintaan dari pengguna akan diarahkan ke fungsi yang telah ditentukan. Sebagai contoh, ketika pengguna mengakses halaman tertentu, sistem secara otomatis menjalankan proses yang sesuai dengan kebutuhan halaman tersebut. Dengan adanya routing, alur navigasi aplikasi menjadi lebih terstruktur, sehingga memudahkan pengelolaan fitur dan alur sistem.

2) Request dan Response (Siklus Interaksi Data)

Dalam proses interaksi antara pengguna dan sistem, Flask mengatur pertukaran data melalui mekanisme *request* dan *response*. Objek *request* berfungsi untuk menerima setiap masukan yang diberikan oleh pengguna, misalnya data hasil pengisian kuesioner mahasiswa. Setelah itu, data tersebut diproses oleh bagian backend sesuai dengan logika sistem yang telah dirancang. Hasil pengolahan tersebut kemudian dikirimkan kembali kepada pengguna dalam bentuk objek *response* yang ditampilkan melalui peramban (*browser*). Mekanisme ini memungkinkan terjadinya komunikasi dua arah yang berjalan secara langsung, sehingga setiap input pengguna dapat segera diproses dan ditampilkan hasilnya tanpa perlu waktu yang lama.

Dalam pengembangan sistem informasi cerdas, Flask memiliki peran penting sebagai media untuk *deployment* model *machine learning*. Secara konsep, model yang telah selesai dilatih sebenarnya masih bersifat pasif, karena hanya tersimpan di dalam memori dan belum dapat digunakan secara langsung oleh pengguna. Agar model tersebut dapat dimanfaatkan secara luas, diperlukan proses penyimpanan

permanen yang dikenal sebagai *persistensi model*. Dalam hal ini, Flask berfungsi sebagai penghubung yang memungkinkan model tersebut diakses melalui jaringan internet. Flask bekerja sebagai lapisan antarmuka yang menyediakan jalur komunikasi berbasis HTTP, sehingga model yang telah disimpan dapat dipanggil kembali dan digunakan untuk melakukan proses prediksi secara dinamis sesuai dengan input dari pengguna (Grinberg, 2018).

Secara teknis, terdapat beberapa format yang umum digunakan untuk menyimpan model dalam ekosistem Python, di antaranya adalah Pickle (.pkl) dan H5 (.h5). Kedua format ini memiliki karakteristik yang berbeda, baik dari segi cara penyimpanan, fleksibilitas, maupun efisiensi penggunaannya dalam system:

1) Format Pickle (.pkl)

   Pickle merupakan metode standar dalam Python yang digunakan untuk menyimpan objek ke dalam bentuk biner. Proses ini dikenal sebagai serialisasi, yaitu mengubah objek Python menjadi aliran data yang dapat disimpan dalam file. Namun, karena sifatnya yang sangat bergantung pada Python, file dengan format Pickle hanya dapat digunakan kembali dalam lingkungan Python yang serupa. Dalam konteks sistem cerdas, format ini sering digunakan untuk menyimpan kondisi model secara lengkap, sehingga ketika dipanggil kembali oleh Flask, model dapat berjalan dengan kondisi yang sama seperti saat pertama kali disimpan.

2) Format H5 (.h5)

   Berbeda dengan format Pickle, H5 (*Hierarchical Data Format versi 5*) merupakan bentuk penyimpanan data dalam file biner yang tersusun seperti

sistem folder atau direktori. H5 menyimpan data seperti *dataset* dan bobot model dalam struktur bertingkat yang rapi dan terorganisir di dalam satu file. Selain itu, format H5 tidak bergantung pada bahasa pemrograman tertentu (*language-independent*), sehingga file yang disimpan dapat dibaca dan digunakan kembali oleh berbagai bahasa pemrograman, tidak hanya Python.

7. MySql

MySQL secara teori termasuk ke dalam sistem manajemen basis data relasional atau *Relational Database Management System (RDBMS)* yang menggunakan bahasa SQL (*Structured Query Language*) untuk mengelola data. Menurut (Nixon, 2018), konsep relasional pada MySQL berarti data tidak disimpan dalam satu tempat besar, tetapi dipisahkan ke dalam beberapa tabel yang saling terhubung satu sama lain. Cara ini membuat data lebih teratur, mengurangi duplikasi, dan menjaga konsistensi informasi di dalam sistem. Karena bersifat *open source*, MySQL banyak digunakan dalam pengembangan aplikasi web karena mampu menangani banyak proses data secara bersamaan dengan stabil dan andal.

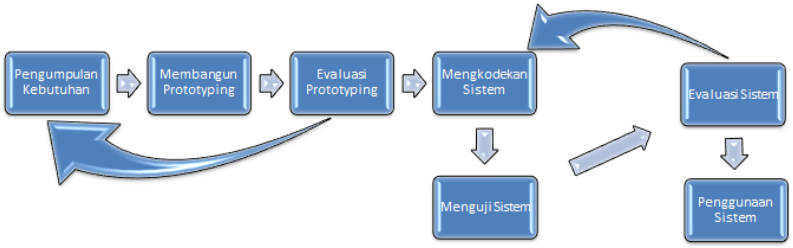
Struktur data pada MySQL disusun secara bertingkat, mulai dari database, tabel, baris, hingga kolom. Setiap tabel biasanya memiliki *primary key* yang berfungsi sebagai identitas unik untuk membedakan setiap data. Sementara itu, hubungan antar tabel dibuat menggunakan *foreign key*. Menurut (Nixon, 2018), struktur ini membuat pengelolaan data menjadi lebih terorganisir meskipun datanya kompleks

dan saling berkaitan, sehingga proses pencarian dan pengambilan data dapat dilakukan dengan lebih cepat dan tepat.

Dalam konteks *machine learning*, MySQL berperan sebagai tempat penyimpanan data utama, baik data mentah maupun data yang sudah diproses. Menurut (Othman & Yasin, 2025), integrasi kecerdasan buatan dengan sistem basis data bertujuan untuk meningkatkan efisiensi pengelolaan data, termasuk dalam hal kecepatan akses dan keamanan. Data yang tersimpan di MySQL kemudian dapat digunakan oleh bahasa pemrograman seperti Python untuk proses analisis dan pelatihan model. Kualitas hasil sistem cerdas sangat dipengaruhi oleh kualitas data yang dikelola di dalam database tersebut, karena data inilah yang menjadi dasar pembentukan pola oleh algoritma machine learning.

8. Prototyping Model

Prototyping Model merupakan pendekatan dalam pengembangan sistem informasi yang menekankan pada pembuatan versi awal sistem atau purwarupa sebagai gambaran dari sistem yang akan dibangun secara penuh. Menurut (Gunawan & Nasution, 2025), model ini dirancang untuk membantu menjembatani perbedaan pemahaman antara pengembang dan pengguna melalui proses komunikasi yang dilakukan secara berulang. Pendekatan ini sangat cocok digunakan ketika kebutuhan sistem belum sepenuhnya jelas di awal pengembangan, karena memungkinkan adanya penyesuaian dan perbaikan secara bertahap berdasarkan hasil evaluasi pengguna. (Haruna et al., 2024)




Gambar 2. 6 Tahapan prototyping model (Arviansyah, 2021)

Berdasarkan rujukan (Gunawan & Nasution, 2025), Prototyping Model merupakan metode pengembangan sistem yang dilakukan secara bertahap dan berulang dengan melibatkan pengguna secara langsung. Model ini bertujuan untuk memastikan bahwa sistem yang dibangun benar-benar sesuai dengan kebutuhan pengguna melalui proses evaluasi yang terus-menerus. Adapun tahapan dalam Prototyping Model adalah sebagai berikut:

1. Pengumpulan Kebutuhan

   Tahap ini merupakan langkah awal dalam pengembangan sistem, di mana pengembang melakukan komunikasi intensif dengan pengguna untuk menggali seluruh kebutuhan sistem. Proses ini mencakup penentuan tujuan utama sistem, identifikasi masalah yang ingin diselesaikan, serta pendataan fitur-fitur yang dibutuhkan. Hasil dari tahap ini menjadi dasar utama dalam proses perancangan sistem agar tidak terjadi perbedaan persepsi antara pengembang dan pengguna.

2. Membangun Prototipe

   Pada tahap ini, pengembang mulai membuat versi awal sistem atau

purwarupa yang masih bersifat sederhana. Prototipe ini tidak langsung dibuat sempurna, tetapi hanya menampilkan gambaran umum dari sistem, seperti alur proses dan tampilan antarmuka. Tujuannya adalah agar pengguna dapat melihat dan memahami bagaimana sistem akan bekerja sebelum dikembangkan lebih lanjut. Dengan cara ini, kesalahan desain dapat diminimalkan sejak awal.

3. Evaluasi Prototipe

   Setelah prototipe selesai dibuat, pengguna akan mencoba dan memberikan penilaian terhadap sistem tersebut. Pada tahap ini, pengguna dapat memberikan masukan, kritik, atau permintaan perubahan jika terdapat fitur yang belum sesuai kebutuhan. Jika masih terdapat ketidaksesuaian, maka pengembang akan kembali memperbaiki prototipe hingga mencapai kesepakatan antara kedua pihak. Proses ini bersifat iteratif atau dapat berulang (Muzayyana Agustin et al., 2024).

4. Pengkodean Sistem

   Jika prototipe telah disetujui, maka tahap selanjutnya adalah mengubah rancangan tersebut menjadi sistem yang sesungguhnya melalui proses pengkodean. Pada tahap ini, pengembang mulai membangun logika sistem secara penuh, termasuk pengolahan data, backend sistem, serta integrasi dengan database. Tahap ini menjadi inti dari implementasi sistem yang sebenarnya.

5. Pengujian Sistem

   Setelah sistem selesai dibangun, dilakukan pengujian untuk memastikan

seluruh fungsi berjalan dengan baik. Pengujian ini bertujuan untuk mendeteksi kesalahan atau bug, serta memastikan bahwa setiap fitur bekerja sesuai dengan kebutuhan yang telah ditentukan pada tahap awal. Selain itu, pengujian juga dilakukan untuk memastikan kestabilan sistem dalam berbagai kondisi penggunaan.

6. Evaluasi Sistem

   Pada tahap ini, sistem yang telah diuji akan dievaluasi kembali oleh pengguna secara menyeluruh. Evaluasi ini bertujuan untuk memastikan bahwa sistem sudah benar-benar sesuai dengan kebutuhan dan layak digunakan. Jika masih ditemukan kekurangan, maka pengembang dapat melakukan perbaikan sebelum sistem diterapkan secara penuh.

7. Penggunaan Sistem

   Tahap terakhir adalah implementasi sistem ke dalam lingkungan nyata, di mana sistem mulai digunakan oleh pengguna dalam aktivitas sehari-hari. Pada tahap ini, sistem sudah dianggap siap digunakan, namun tetap memerlukan pemeliharaan dan pengawasan agar tetap berjalan optimal serta dapat diperbaiki jika ditemukan masalah di kemudian hari.

## C. KEASLIAN PENELITIAN

**Sistem Rekomendasi Minat Akademik Mahasiswa Berbasis Machine Learning Menggunakan Metode K-Means dan Decision Tree**

*Table 2. 1 Matriks Literatur Review dan Posisi Penelitian*

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|---|---|---|---|---|---|---|
| 1 | Penerapan Algoritma K-Means untuk Mengidentifikasi Minat dan Bakat Siswa Sekolah Dasar | Sonianto, Hartono. *Jurnal Sienna: Jurnal Ilmu Komputer dan Teknologi Informasi*, Volume 6, Nomor 1, Tahun 2025 (Sonianto & Hartono, 2025) | Penelitian ini bertujuan untuk membantu tenaga pendidik dalam mengenali minat dan bakat siswa sejak dini dengan memanfaatkan pengolahan data, sehingga proses identifikasi tidak hanya bergantung pada penilaian subjektif guru, tetapi juga didukung oleh | Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan siswa ke dalam beberapa cluster berdasarkan kemiripan atribut seperti nilai, aktivitas, dan kecenderungan tertentu. Setiap cluster | Penelitian ini hanya berfokus pada proses pengelompokan data tanpa adanya tahap lanjutan seperti klasifikasi atau prediksi. Selain itu, hasil yang diperoleh masih bersifat deskriptif sehingga belum dapat digunakan untuk memprediksi minat | Penelitian ini hanya menggunakan K-Means untuk mengelompokkan data siswa berdasarkan kemiripan atribut sehingga output yang dihasilkan berupa cluster tanpa interpretasi lanjutan. Sementara itu, penelitian yang diusulkan tidak hanya melakukan clustering menggunakan K-Means, tetapi juga |

| *Table 2. 1 Matriks Literatur Review dan Posisi Penelitian* | | | | | | |
|---|---|---|---|---|---|---|
| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
| | | | hasil analisis data yang lebih terstruktur. | merepresentasikan kelompok minat yang berbeda sehingga dapat memberikan gambaran awal mengenai potensi siswa. | siswa di masa mendatang. | melanjutkan hasil cluster sebagai input untuk algoritma Random Forest. Dengan alur ini, sistem mampu melakukan prediksi minat siswa baru berdasarkan pola dari cluster, sehingga perbedaan terletak pada adanya tahap klasifikasi lanjutan dan perubahan output dari deskriptif menjadi prediktif. |
| 2 | Prediksi Minat Siswa Berdasarkan Tes OCEAN dan Aptitude | Muhamad Sidik, Andri Kurniawan, Fariz Faqih, Farizi Ilham. *Jurnal Sistem Pendukung* | Penelitian ini bertujuan untuk membantu siswa dalam memahami minatnya dengan memanfaatkan data | Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan minat siswa dengan | Penelitian ini langsung melakukan proses klasifikasi tanpa melalui tahap analisis awal seperti clustering, sehingga | Penelitian ini menggunakan Random Forest secara langsung untuk klasifikasi tanpa melakukan analisis pola data terlebih dahulu. |

| *Table 2. 1 Matriks Literatur Review dan Posisi Penelitian* | | | | | | |
|---|---|---|---|---|---|---|
| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
| | Menggunakan Random Forest | *Keputusan dan Kecerdasan Buatan*, Volume 3, Nomor 1, Tahun 2025 (Sidik et al., 2025) | hasil tes kepribadian OCEAN dan aptitude test yang diolah menggunakan algoritma Random Forest. | cukup baik berdasarkan kombinasi data kepribadian dan kemampuan, sehingga sistem dapat memberikan hasil prediksi yang konsisten. | model tidak mempertimbangkan struktur awal data dan hubungan antar kelompok data. | Sedangkan penelitian yang diusulkan menambahkan tahap clustering menggunakan K-Means sebelum proses klasifikasi. Dengan pendekatan ini, data terlebih dahulu dikelompokkan sehingga pola distribusi data dapat dikenali, kemudian hasilnya digunakan dalam proses klasifikasi, sehingga alur sistem menjadi lebih terstruktur dan berpotensi meningkatkan akurasi prediksi. |

| *Table 2. 1 Matriks Literatur Review dan Posisi Penelitian* | | | | | | |
|---|---|---|---|---|---|---|
| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
| 3 | Implementasi Aplikasi Web Pemilihan Kelas Berdasarkan Minat Menggunakan K-Means Clustering | Clarenza Dixie Rose, Bernadus Anggo Seno Aji, Farah Zakiyah Rahmanti. *Jurnal Teknologi Informasi dan Komunikasi*, Volume 9, Nomor 1, Tahun 2025 (Rose et al., 2024) | Penelitian ini bertujuan untuk mengembangkan sistem berbasis web yang dapat membantu pihak sekolah dalam menentukan pembagian kelas siswa berdasarkan minat secara otomatis. | Hasil penelitian menunjukkan bahwa sistem yang dibangun mampu mengelompokkan siswa menggunakan algoritma K-Means sehingga proses pembagian kelas menjadi lebih cepat, terstruktur, dan efisien. | Sistem yang dikembangkan hanya berfungsi untuk pengelompokan siswa tanpa adanya fitur analisis lanjutan seperti prediksi atau evaluasi terhadap minat siswa di masa depan. | Penelitian ini hanya memanfaatkan K-Means dalam sistem web untuk menghasilkan pembagian kelompok siswa berdasarkan minat. Sementara itu, penelitian yang diusulkan tidak hanya mengimplementasikan sistem clustering, tetapi juga menambahkan fitur analisis dan prediksi menggunakan Random Forest. Dengan demikian, sistem yang diusulkan tidak hanya menghasilkan output berupa kelompok, tetapi |

| *Table 2. 1 Matriks Literatur Review dan Posisi Penelitian* | | | | | | |
|---|---|---|---|---|---|---|
| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
| | | | | | | juga mampu memberikan prediksi minat siswa sehingga fungsi sistem menjadi lebih komprehensif. |
| 4 | Sistem Pendukung Keputusan Identifikasi Minat dan Bakat Siswa Sekolah Menengah Atas | Desy Rahmallia. *Jurnal Pendidikan Indonesia*, Volume 4, Nomor 3, Tahun 2024 (Rahmallia, 2024) | Penelitian ini bertujuan untuk meningkatkan ketepatan dalam menentukan minat dan bakat siswa dengan memanfaatkan algoritma Decision Tree dan Random Forest dalam sistem pendukung keputusan. | Hasil penelitian menunjukkan bahwa penggunaan algoritma Decision Tree dan Random Forest mampu memberikan hasil klasifikasi yang lebih akurat dibandingkan metode manual, sehingga dapat membantu pengambilan | Penelitian ini hanya menggunakan pendekatan klasifikasi tanpa melakukan analisis awal terhadap pola data menggunakan metode clustering, sehingga belum mampu menangkap struktur kelompok data secara menyeluruh. | Penelitian ini langsung menggunakan Decision Tree dan Random Forest untuk melakukan klasifikasi minat siswa tanpa tahap clustering. Sedangkan penelitian yang diusulkan menambahkan proses clustering menggunakan K-Means sebelum klasifikasi. Dengan demikian, sistem yang diusulkan mampu memahami pola |

| *Table 2. 1 Matriks Literatur Review dan Posisi Penelitian* | | | | | | |
|---|---|---|---|---|---|---|
| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
| | | | | keputusan secara lebih objektif. | | kelompok data terlebih dahulu sebelum melakukan klasifikasi, sehingga analisis menjadi lebih mendalam dan tidak hanya berbasis pada model klasifikasi saja. |
| 5 | Implementation of Decision Tree for Student Interest Analysis | Muhammad Haekal, Irvan. *Journal of Artificial Intelligence Engineering*, Volume 1, Nomor 1, Tahun 2025 (Muhammad & Irvan, 2025) | Penelitian ini bertujuan untuk menganalisis minat siswa terhadap mata pelajaran tertentu menggunakan algoritma Decision Tree sebagai metode klasifikasi. | Hasil penelitian menunjukkan bahwa Decision Tree mampu menghasilkan aturan keputusan yang mudah dipahami dan dapat digunakan untuk menentukan minat siswa berdasarkan data yang tersedia. | Penelitian ini hanya menggunakan satu algoritma tanpa melakukan perbandingan atau kombinasi dengan algoritma lain, sehingga belum diketahui apakah metode tersebut merupakan yang paling optimal. | Penelitian ini menggunakan Decision Tree sebagai metode tunggal untuk klasifikasi sehingga hanya menghasilkan aturan keputusan sederhana. Sedangkan penelitian yang diusulkan menggunakan kombinasi algoritma, yaitu K-Means untuk |

*Table 2. 1 Matriks Literatur Review dan Posisi Penelitian*

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|---|---|---|---|---|---|---|
| | | | | | | clustering dan Random Forest untuk klasifikasi. Dengan kombinasi ini, sistem tidak hanya menghasilkan aturan, tetapi juga mampu menangkap pola data yang lebih kompleks dan meningkatkan kualitas hasil analisis. |
| 6 | Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms | Mustafa Yağcı. *Smart Learning Environments*, Volume 9, Nomor 1, Tahun 2022 (Yağcı, 2022) | Penelitian ini bertujuan untuk memprediksi performa akademik siswa dengan menggunakan berbagai algoritma seperti Random Forest, Support Vector Machine, | Hasil penelitian menunjukkan bahwa model yang dibangun mampu menghasilkan tingkat akurasi yang cukup baik dalam memprediksi performa akademik siswa, dengan | Penelitian ini hanya berfokus pada performa akademik siswa dan tidak menjadikan minat siswa sebagai variabel utama, sehingga aspek minat tidak dianalisis secara khusus. | Penelitian ini menggunakan berbagai algoritma untuk memprediksi performa akademik siswa berdasarkan nilai dan aktivitas belajar. Sedangkan penelitian yang diusulkan berfokus pada analisis minat |

*Table 2. 1 Matriks Literatur Review dan Posisi Penelitian*

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|---|---|---|---|---|---|---|
| | | | Logistic Regression, Naïve Bayes, dan K-Nearest Neighbor berdasarkan data akademik siswa. | Random Forest sebagai salah satu algoritma yang memiliki performa tinggi. | | siswa sebagai variabel utama dengan menggunakan K-Means untuk clustering dan Random Forest untuk klasifikasi. Dengan demikian, perbedaan terletak pada tujuan analisis, variabel yang digunakan, serta output yang dihasilkan. |
| 7 | Career Guidance System Using Decision Tree, Random Forest, and Naïve Bayes Algorithm | Chidi Ukamaka Betrand, Obinna Banner Aliche, Chinwe Gilean Onukwugha, Christopher Ifeanyi Ofoegbu, Douglas Allswell Kelechi, | Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi karir berbasis web yang dapat membantu pengguna menentukan pilihan karir berdasarkan minat, | Hasil penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi karir yang sesuai dengan profil pengguna, dengan algoritma Random Forest | Penelitian ini berfokus pada tahap akhir yaitu pemberian rekomendasi karir dan tidak membahas secara rinci proses awal identifikasi minat siswa. | Penelitian ini menggunakan beberapa algoritma klasifikasi untuk memberikan rekomendasi karir berdasarkan data minat yang sudah tersedia. Sedangkan penelitian yang diusulkan berfokus |

| *Table 2. 1 Matriks Literatur Review dan Posisi Penelitian* | | | | | | |
|---|---|---|---|---|---|---|
| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
| | | Ihechiluru Chinwe Ugbor, Nneka Martina Oragba. *International Journal of Science, Technology and Society*, Volume 13, Nomor 2, Tahun 2025 (Betrand et al., 2025) | kemampuan, dan profil individu. | memberikan hasil yang paling optimal dibandingkan algoritma lainnya. | | pada tahap awal, yaitu proses identifikasi dan analisis minat siswa menggunakan K-Means dan Random Forest. Dengan demikian, perbedaan utama terletak pada tahapan sistem, yaitu penelitian ini berada pada tahap rekomendasi akhir, sedangkan penelitian yang diusulkan berada pada tahap identifikasi awal minat siswa. |