BAB 2
LANDASAN TEORI

2.1 Intelligent Tutoring System (ITS)

Intelligent Tutoring System adalah sistem pendidikan adaptif yang menggunakan teknik kecerdasan buatan untuk memberikan pengajaran yang dipersonalisasi [22]. Secara teknis, *Intelligent Tutoring System* adalah program komputer yang bertujuan untuk memberikan instruksi atau umpan balik langsung dan sesuai kebutuhan kepada peserta didik dengan menggunakan model komputasi berdasarkan penelitian di bidang kecerdasan buatan dan ilmu kognitif [23]. Sistem pembelajaran pintar ini dibuat untuk melakukan klasifikasi gaya belajar siswa yang didapatkan dari pengalaman belajar melalui latihan soal yang diberikan.

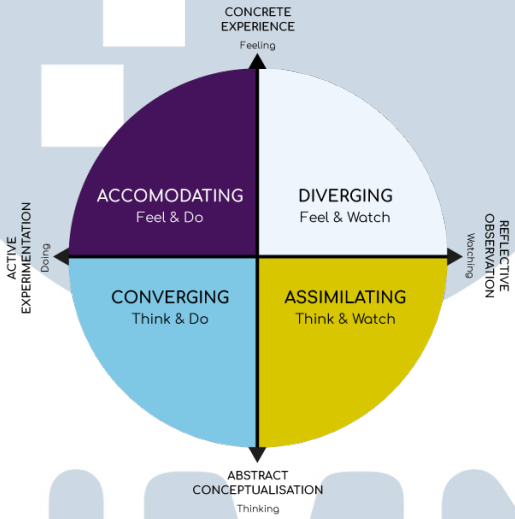
2.2 Tantangan Pendidik dalam Mengakomodasi Gaya Belajar Siswa

Pemahaman terhadap gaya belajar siswa merupakan hal penting dalam proses pembelajaran. Namun, implementasi strategi pengajaran yang responsif terhadap variasi gaya belajar siswa sering menjadi tantangan dalam praktik kelas sehari-hari. Sebuah penelitian menunjukkan bahwa guru menghadapi kendala signifikan dalam merespons kebutuhan belajar individual siswa, termasuk faktor keterbatasan waktu dan dukungan profesional dalam mengimplementasikan pembelajaran inklusif yang efektif [24]. Selain itu, hubungan antara gaya belajar siswa dan gaya pengajaran guru menunjukkan bahwa meskipun menyesuaikan strategi pembelajaran dengan berbagai gaya belajar dapat meningkatkan hasil belajar, proses menyelaraskan keduanya bukanlah hal yang sederhana dalam konteks pendidikan nyata [25]. Tantangan ini juga dirasakan dalam penerapan *differentiated instruction*, di mana guru harus merencanakan dan menyiapkan materi yang adaptif untuk beragam kebutuhan belajar siswa, yang sering kali menuntut waktu dan sumber daya yang tidak sedikit [26].

2.3 Kolb Learning Style

Model gaya belajar Kolb adalah model gaya belajar yang paling banyak diterima dan telah menerima dukungan empiris [13]. Model gaya belajar Kolb didasarkan pada teori *Experiential Learning*. Teori *Experiental Learning*

memandang pembelajaran sebagai proses adaptif yang bersifat holistik, di mana individu membangun pengetahuan melalui pengalaman langsung dan refleksi terhadap pengalaman tersebut [16]. Hal ini berbeda dengan pendekatan behavioristik maupun metode pendidikan tradisional yang berorientasi pada transfer pengetahuan, teori *Experiental Learning* menempatkan pengalaman sebagai pusat proses pembelajaran manusia. Kolb yang berdasar pada teori *Experiential Learning* mengemukakan bahwa terdapat empat tahapan dalam proses belajar berlangsung, yaitu Concrete Experience (CE), Reflective Observation (RO), Abstract Conceptualization (AC), dan Active Experimentation (AE) [12]. Gambar 2.1 menunjukkan siklus pembelajaran *experiential learning* pada model Kolb.



Gambar 2.1. Siklus Pembelajaran Model Kolb [1]

2.3.1 Concrete Experience

Concrete Experience merupakan tahap awal dalam siklus *Experiential Learning*. Dalam *Concrete Experience*, individu sepenuhnya dan secara terbuka terlibat dalam pengalaman belajar dan tidak membiarkan prasangka memengaruhi persepsinya [27]. Sehingga dapat dikatakan bahwa pembelajaran ini terjadi melalui keterlibatan aktif terhadap peristiwa tertentu tanpa terlebih dahulu melakukan analisis konseptual yang mendalam. Individu yang dominan pada dimensi ini cenderung mengandalkan perasaan, intuisi, serta keterlibatan langsung dalam memahami suatu fenomena [16, 28].

2.3.2 Reflective Observation

Pada tahap *Reflective Observation*, individu diharuskan untuk merefleksikan dan mengamati pengalamannya dari berbagai perspektif yang berbeda [27]. Proses belajar berlangsung melalui pengamatan yang cermat, analisis, serta pertimbangan dari berbagai sudut pandang. Individu dengan kecenderungan RO lebih menekankan pada proses memahami makna pengalaman sebelum mengambil tindakan lebih lanjut. Refleksi ini berfungsi sebagai jembatan antara pengalaman konkret dan pembentukan konsep abstrak dalam siklus pembelajaran [16, 29].

2.3.3 Abstract Conceptualization

Pada tahap *Abstract Conceptualization* merupakan tahap di mana individu menguraikan konsep-konsep yang diintegrasikan melalui pengalaman sebelumnya menjadi penjelasan yang logis dan masuk akal [27]. Pembelajaran pada tahap ini berorientasi pada pemikiran logis, analitis, dan sistematis. Individu yang memiliki kecenderungan AC lebih mengutamakan pemahaman konseptual dibandingkan pengalaman langsung sehingga lebih cenderung menyusun model, kerangka teori, atau prinsip abstrak untuk menjelaskan fenomena yang diamati [16, 28].

2.3.4 Active Experimentation

Tahap *Active Experimentation* merupakan tahap individu mampu untuk menggunakan dan menerapkan teori untuk mengambil keputusan dan memecahkan masalah [27]. Pembelajaran terwujud melalui tindakan, pengambilan keputusan, serta percobaan aktif dalam kehidupan. Individu dengan kecenderungan AE lebih fokus pada penerapan praktis dan pemecahan masalah secara langsung. Proses ini menghasilkan pengalaman baru yang kembali memasuki tahap *Concrete Experience* sehingga membentuk siklus pembelajaran berkelanjutan [16, 29].

Berdasarkan kombinasi dominan dari dua tahapan dalam siklus tersebut, Kolb mengklasifikasikan gaya belajar ke dalam empat kategori, yaitu diverger, assimilator, converger, dan accommodator [14, 12, 13].

A Diverger

Tipe gaya belajar ini menggambarkan individu yang belajar melalui pengalaman konkret [13]. Tipe gaya belajar *diverger* merupakan kombinasi

antara *Concrete Experience* dan *Reflective Observation* [15, 13]. Individu dengan tipe ini cenderung memiliki kemampuan dalam melihat suatu situasi dari berbagai sudut pandang serta mengembangkan ide-ide kreatif sehingga lebih mengandalkan pengalaman konkret dan refleksi mendalam sebelum membuat keputusan. Tipe gaya belajar *diverger* sering menunjukkan kekuatan dalam imajinasi, sensitivitas terhadap perasaan, serta kemampuan mengidentifikasi makna dari suatu pengalaman [16, 28].

B Assimilator

Tipe gaya belajar *assimilator* merupakan kombinasi antara *Abstract Conceptualization* dan *Reflective Observation* yang memungkinkan individu pada gaya belajar ini lebih menekankan pemahaman logis [13]. Individu pada gaya belajar ini cenderung tertarik pada teori, model, serta struktur sistematis dalam memahami suatu fenomena. Tipe gaya belajar *assimilator* unggul dalam mengorganisasi informasi secara terstruktur serta membangun kerangka konseptual yang koheren [16, 28]

C Converger

Tipe gaya belajar *converger* adalah kombinasi antara *Abstract Conceptualization* dan *Active Experimentation*. Individu dengan tipe ini menemukan kegunaan praktis dari ide dan teori yang telah dipelajari sehingga memungkinkan untuk bisa memecahkan masalah baru dengan solusi dari masalah-masalah sebelumnya [13]. Tipe gaya belajar *converger* menunjukkan kemampuan kuat dalam pemecahan masalah serta pengujian ide melalui eksperimen aktif [16, 28].

D Accomodator

Tipe gaya belajar *accomodator* merupakan kombinasi antara *Concrete Experience* dan *Active Experimentation*. Individu dengan tipe ini cenderung belajar melalui pengalaman baru serta secara aktif melaksanakan rencana yang melibatkan pengalaman dan tantangan baru [13]. Tipe *accomodator* lebih memilih pendekatan *trial-and-error* dalam menyelesaikan permasalahan serta belajar melalui keterlibatan langsung dibandingkan analisis teoritis yang mendalam [16, 28].

2.4 Data Collection

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui proses interaksi pengguna dengan sistem identifikasi gaya belajar Kolb yang telah dikembangkan. Data primer merupakan data yang diperoleh secara langsung dari sumber penelitian untuk menjawab tujuan penelitian tertentu [30].

Pengumpulan data dilakukan dengan mencatat aktivitas pengguna selama mengerjakan instrumen identifikasi gaya belajar Kolb. Data yang dikumpulkan meliputi pilihan jawaban pengguna pada setiap soal serta waktu pengerjaan yang dibutuhkan untuk menyelesaikan setiap soal. Data tersebut digunakan untuk merepresentasikan karakteristik perilaku pengguna selama proses pengerjaan instrumen, yang selanjutnya akan diolah melalui tahap *feature engineering* untuk menghasilkan fitur numerik sebagai masukan pada proses klasifikasi.

Pada penelitian ini, data interaksi pengguna digunakan sebagai dasar dalam pembentukan dataset klasifikasi gaya belajar Kolb. Dataset yang diperoleh kemudian diproses untuk menghasilkan fitur berupa proporsi jawaban, transformasi logaritmik waktu pengerjaan, dan indeks efisiensi. Selanjutnya, fitur-fitur tersebut digunakan untuk melatih dan mengevaluasi model klasifikasi menggunakan algoritma *Gaussian Naive Bayes*.

2.5 Machine Learning

Machine Learning adalah cabang kecerdasan buatan yang berkaitan dengan pengembangan algoritma dan model yang dapat secara otomatis mempelajari pola dari data dan melakukan tugas tanpa instruksi eksplisit [31]. Teknik dalam *Machine Learning* terbagi menjadi beberapa kategori seperti klasifikasi, regresi, dan pengelompokan [32]. Definisi *machine learning* secara formal dikatakan bahwa suatu program komputer belajar dari pengalaman E sehubungan dengan beberapa kelas tugas T dan ukuran kinerja P apabila performanya pada tugas T, yang diukur dengan P, meningkat seiring dengan pengalaman E [33]. Pendekatan *machine learning* bersifat *data-driven*, sehingga performa algoritma *machine learning* bergantung pada kualitas dan kuantitas data yang tersedia yang relevan dengan tugas yang sedang dikerjakan [34]. Hal ini berbeda dengan pendekatan *rule-based system* yang mengandalkan aturan yang telah ditentukan sebelumnya dan kondisi logis yang dapat dieksekusi dengan cepat [35].

Secara umum, machine learning dapat diklasifikasikan menjadi tiga kategori

utama, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning* [31, 34]. Pada *supervised learning*, model dilatih menggunakan data yang telah memiliki label sehingga sistem mempelajari hubungan fungsional antara variabel masukan (*input features*) dan variabel keluaran (*label*) [31]. Pendekatan ini bertujuan untuk mempelajari suatu fungsi pemetaan dari input ke output berdasarkan data yang tersedia, di mana algoritma digunakan untuk mengestimasi parameter fungsi tersebut agar mampu melakukan prediksi terhadap data baru [34]. Sebaliknya, *unsupervised learning* merupakan pendekatan *machine learning* yang tidak menggunakan label dalam proses pelatihannya. Tujuan utama dari unsupervised learning adalah mengidentifikasi pola, struktur, atau hubungan tersembunyi dalam data tanpa adanya informasi mengenai keluaran yang benar [31]. Sementara itu, *reinforcement learning* merupakan pendekatan *machine learning* yang meniru proses pembelajaran manusia melalui interaksi dengan lingkungan.

Dalam permasalahan klasifikasi, model berupaya mempelajari suatu fungsi pemetaan:

$$f : X \rightarrow Y \tag{2.1}$$

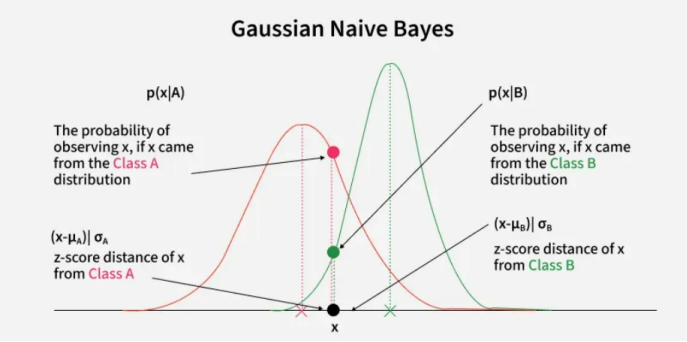
di mana X merepresentasikan vektor fitur dan Y merepresentasikan kelas diskrit [36]. Dalam penelitian klasifikasi gaya belajar Kolb, pendekatan *supervised learning* digunakan karena proses pelatihan model membutuhkan data masukan (*input features*) dan label sebagai target klasifikasi. Variabel masukan X direpresentasikan dalam bentuk vektor fitur yang diperoleh dari hasil pengolahan data interaksi pengguna selama mengerjakan instrumen gaya belajar Kolb. Fitur tersebut meliputi proporsi jawaban pada setiap dimensi gaya belajar, transformasi logaritmik waktu pengerjaan, serta indeks efisiensi. Sementara itu, variabel keluaran Y merupakan label kelas gaya belajar Kolb yang terdiri atas *Diverger*, *Assimilator*, *Converger*, dan *Accommodator*. Label tersebut diperoleh melalui proses pembentukan label berdasarkan kombinasi kecenderungan dimensi gaya belajar Kolb, sehingga dapat digunakan sebagai *ground truth* pada proses pelatihan model klasifikasi.

2.6 Educational Data Mining

Educational Data Mining (EDM) merupakan bidang penelitian yang berfokus pada pengembangan metode untuk mengeksplorasi data yang berasal

dari lingkungan pendidikan guna memahami karakteristik peserta didik, proses pembelajaran, serta kondisi yang memengaruhi hasil belajar [37]. EDM menggunakan data dari interaksi yang terjadi sepanjang proses belajar. Data tersebut bisa mencakup berbagai aktivitas seperti mengerjakan soal, cara berpindah di dalam sistem, lama waktu yang digunakan untuk mengerjakan, berapa kali percobaan dilakukan, hingga sejarah interaksi pengguna pada platform pembelajaran digital [38]. Salah satu tujuan utama EDM adalah membuat model siswa, yaitu gambaran berbentuk komputer yang menampilkan sifat, kemampuan, tindakan, atau selera belajar siswa berdasarkan informasi yang diperoleh selama proses belajar [37, 39]. Dalam lingkungan belajar menggunakan komputer, data tentang cara siswa berperilaku sering digunakan sebagai informasi untuk memahami bagaimana karakteristik siswa itu berupa. Beberapa penelitian menyebutkan bahwa cara seseorang menjawab soal, kemampuan menyelesaikan tugas, serta lama waktu yang digunakan dalam mengerjakan tugas bisa digunakan untuk mengetahui strategi belajar, tingkat pemahaman materi, dan juga sifat-sifat pribadi seseorang selama belajar [37, 38]. Waktu respons juga merupakan salah satu jenis data proses yang sering digunakan dalam EDM. Data waktu bisa memberikan informasi tambahan mengenai cara siswa memahami materi, membuat keputusan, dan menyelesaikan tugas belajar [40, 41].

2.7 Gaussian Naive Bayes



Gambar 2.2. Kurva Distribusi Gaussian Naive Bayes
Sumber: GeeksforGeeks [42]

Gaussian Naive Bayes adalah salah satu jenis algoritma Naive Bayes yang dibuat khusus untuk mengolah data berupa angka yang berkelanjutan. Algoritma ini menggunakan Teorema Bayes dan pendekatan probabilistik untuk mengetahui

seberapa besar kemungkinan suatu data masuk ke dalam suatu kelas tertentu berdasarkan fitur yang dimiliki oleh data tersebut [36, 43]. Berbeda dengan Naive Bayes yang digunakan untuk data kategorikal, Gaussian Naive Bayes menganggap bahwa setiap fitur memiliki bentuk distribusi normal (Gaussian) di setiap kelas.

Gambar 2.2 menampilkan bentuk distribusi Gaussian yang digunakan sebagai dasar dalam menghitung probabilitas pada model Gaussian Naive Bayes. Distribusi ini memiliki bentuk kurva lonceng yang simetris di sekitar nilai rata-rata (μ). Asumsi distribusi Gaussian digunakan karena semua fitur yang digunakan berupa data numerik kontinu, yaitu proporsi jawaban, log waktu pengerjaan, dan indeks efisiensi. Selain memiliki cara pelatihan yang mudah dan hemat komputasi, Gaussian Naive Bayes juga bisa menyelesaikan masalah klasifikasi dengan lebih dari dua kelas secara langsung.

2.7.1 Teorema Bayes

Gaussian Naive Bayes merupakan pengembangan dari algoritma Naive Bayes yang didasarkan pada Teorema Bayes. Teorema ini digunakan untuk menghitung probabilitas suatu kelas berdasarkan data yang diamati [36, 43].

Secara matematis, Teorema Bayes dinyatakan sebagai:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \tag{2.2}$$

dengan:

- $P(C|X)$ adalah probabilitas posterior,
- $P(X|C)$ adalah likelihood,
- $P(C)$ adalah probabilitas prior,
- $P(X)$ adalah evidence.

Dalam proses klasifikasi, tujuan utama adalah menentukan probabilitas posterior untuk setiap kelas yang tersedia. Pada penelitian ini, kelas yang digunakan adalah empat kategori gaya belajar Kolb, yaitu *Diverger*, *Assimilator*, *Converger*, dan *Accommodator*.

2.7.2 Asumsi Independensi Kondisional

Gaussian Naive Bayes mengasumsikan bahwa setiap fitur bersifat independen secara kondisional terhadap kelasnya. Asumsi ini dikenal sebagai *conditional independence assumption* [44].

Misalkan terdapat sejumlah fitur:

$$X = (x_1, x_2, \ldots, x_n) \tag{2.3}$$

maka probabilitas gabungan seluruh fitur pada suatu kelas dapat disederhanakan menjadi:

$$P(X|C) = \prod_{i=1}^{n} P(x_i|C) \tag{2.4}$$

Asumsi ini membuat proses perhitungan probabilitas menjadi lebih sederhana dan efisien. Meskipun pada data nyata antar fitur tidak selalu benar-benar independen, berbagai penelitian menunjukkan bahwa Naive Bayes tetap mampu memberikan performa klasifikasi yang baik pada berbagai kasus [44].

2.7.3 Probabilitas Prior

Probabilitas prior menunjukkan peluang awal suatu kelas sebelum mempertimbangkan nilai fitur yang dimiliki oleh data. Dengan menggabungkan Teorema Bayes dan asumsi independensi kondisional, probabilitas posterior suatu kelas dapat dihitung menggunakan:

$$P(C) = \frac{\text{Jumlah data pada kelas } C}{\text{Jumlah seluruh data}} \tag{2.5}$$

Persamaan probabilitas prior tersebut merupakan bagian dari Teorema Bayes yang digunakan untuk merepresentasikan peluang awal suatu kelas sebelum mempertimbangkan informasi fitur [36]. Probabilitas prior digunakan untuk menunjukkan distribusi awal setiap kelas dalam data pelatihan sebelum memperhitungkan karakteristik dari fitur yang dimiliki oleh siswa. Probabilitas prior dalam hal ini menunjukkan peluang awal suatu data responden termasuk ke dalam salah satu kategori gaya belajar Kolb berdasarkan distribusi kelas pada

dataset yang digunakan.

2.7.4 Likelihood Gaussian

Persamaan distribusi Gaussian digunakan untuk memodelkan distribusi fitur kontinu pada masing-masing kelas dan menjadi dasar perhitungan *likelihood* pada *Gaussian Naive Bayes* [36, 43]. Fitur bersifat kontinu sehingga probabilitas kemunculan suatu fitur pada kelas tertentu dihitung menggunakan:

$$P(x_i|C) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \tag{2.6}$$

dengan:

- x_i adalah nilai fitur ke-i,
- μ adalah rata-rata fitur pada kelas C,
- σ^2 adalah variansi fitur pada kelas C.

Persamaan tersebut digunakan untuk menghitung peluang kemunculan setiap fitur pada masing-masing kelas gaya belajar. Penggunaan distribusi Gaussian didasarkan pada keyakinan bahwa nilai-nilai fitur yang berupa bilangan kontinu dalam setiap kelas mengikuti pola distribusi normal, yang bisa dijelaskan dengan dua parameter, yaitu rata-rata (μ) dan variansi (σ^2). Dalam Gaussian Naive Bayes, parameter tersebut dipelajari secara terpisah untuk setiap fitur dan setiap kelas selama proses pembelajaran model [36, 43].

2.7.5 Probabilitas Posterior

Berdasarkan asumsi independensi kondisional pada Naive Bayes, probabilitas posterior suatu kelas dihitung menggunakan persamaan:

$$P(C|X) \propto P(C) \prod_{i=1}^{n} P(x_i|C) \tag{2.7}$$

dengan:

- $P(C)$ adalah probabilitas prior,

- $P(x_i|C)$ adalah likelihood fitur ke-i,
- n adalah jumlah fitur.

Nilai posterior menyatukan informasi awal yang berasal dari probabilitas prior dengan informasi yang didapatkan dari semua fitur yang diamati. Posterior bisa diartikan sebagai tingkat keyakinan model bahwa seorang siswa masuk ke dalam kelas tertentu setelah memperhatikan karakteristik perilaku yang dimilikinya [36]. Probabilitas posterior dalam hal ini dihitung untuk setiap kategori gaya belajar Kolb berdasarkan kombinasi dari proporsi jawaban, durasi waktu mengerjakan, dan indeks efisiensi.

2.7.6 Log Posterior

Pendekatan logaritmik banyak digunakan pada implementasi Naive Bayes untuk menghindari masalah *floating point underflow* yang muncul akibat perkalian sejumlah probabilitas bernilai sangat kecil [45]. Oleh karena itu, perhitungan dilakukan dalam domain logaritmik:

$$\log P(C|X) = \log P(C) + \sum_{i=1}^{n} \log P(x_i|C) \tag{2.8}$$

Pendekatan ini lebih stabil secara numerik dan umum digunakan pada implementasi Gaussian Naive Bayes. Dalam penerapan nyata, nilai kemungkinan yang dihasilkan dari setiap fitur biasanya sangat kecil karena berupa probabilitas sehingga ketika beberapa kemungkinan dikalikan secara bersamaan, hasil perkaliannya bisa menjadi sangat kecil hingga tidak bisa ditampilkan oleh sistem bilangan komputer (*underflow*). Maka, transformasi logaritmik digunakan untuk mengubah operasi perkalian menjadi operasi penjumlahan tanpa mengubah urutan probabilitas yang dihasilkan [45]. Selain membantu menstabilkan perhitungan, pendekatan log posterior juga membuat proses komputasi lebih cepat dan telah digunakan secara umum dalam berbagai algoritma Naive Bayes.

2.7.7 Keputusan Klasifikasi

Tahap akhir adalah memilih kelas dengan nilai probabilitas posterior terbesar menggunakan pendekatan *Maximum A Posteriori* (MAP):

$$\hat{C} = \arg\max_{C_k} P(C_k|X) \tag{2.9}$$

Kelas dengan probabilitas tertinggi ditetapkan sebagai hasil klasifikasi gaya belajar pengguna. Proses pemilihan kelas yang memiliki probabilitas posterior paling besar disebut sebagai pendekatan *Maximum A Posteriori* (MAP). Cara ini digunakan karena tujuan utama dalam klasifikasi adalah menemukan kelas yang paling mungkin menghasilkan data yang diamati [36]. Nilai posterior dihitung untuk keempat kategori gaya belajar yang dikemukakan oleh Kolb.

2.8 Pembentukan Vektor Fitur

Sebelum dilakukan proses klasifikasi menggunakan algoritma Gaussian Naive Bayes, data interaksi pengguna yang diperoleh dari sistem perlu ditransformasikan menjadi vektor fitur numerik. Tahap ini dikenal sebagai *feature engineering*, yaitu proses ekstraksi dan pembentukan fitur yang merepresentasikan karakteristik perilaku pengguna agar dapat digunakan sebagai masukan (*input*) pada model pembelajaran mesin [46].

Data mentah yang diperoleh dari sistem berupa pilihan jawaban pengguna pada setiap soal serta waktu pengerjaan setiap soal. Data tersebut kemudian diolah menjadi tiga kelompok fitur utama, yaitu fitur proporsi jawaban, fitur transformasi waktu pengerjaan, dan fitur indeks efisiensi. Ketiga kelompok fitur tersebut menghasilkan total tiga belas fitur numerik yang digunakan untuk merepresentasikan karakteristik perilaku pengguna selama mengerjakan instrumen gaya belajar Kolb.

2.8.1 Proporsi Jawaban

Proporsi jawaban digunakan untuk menggambarkan tingkat kecenderungan pengguna terhadap masing-masing dimensi gaya belajar Kolb, yaitu *Concrete Experience* (CE), *Reflective Observation* (RO), *Abstract Conceptualization* (AC), dan *Active Experimentation* (AE).

Proporsi pada setiap dimensi dihitung berdasarkan perbandingan antara jumlah pilihan jawaban yang merepresentasikan suatu dimensi terhadap jumlah seluruh jawaban yang diberikan oleh pengguna. Perhitungan proporsi dinyatakan menggunakan persamaan:

$$\text{proporsi}_i = \frac{n_i}{N} \tag{2.10}$$

dengan:

- n_i adalah jumlah jawaban pengguna pada dimensi ke-i,
- N adalah jumlah keseluruhan soal yang dijawab oleh pengguna.

Nilai proporsi berada pada rentang 0 hingga 1 dan menunjukkan tingkat dominasi suatu dimensi gaya belajar pada seorang pengguna. Semakin tinggi nilai proporsi pada suatu dimensi, semakin besar kecenderungan pengguna terhadap karakteristik gaya belajar yang direpresentasikan oleh dimensi tersebut.

Berdasarkan proses tersebut, dibentuk empat fitur proporsi, yaitu *proporsi_CE*, *proporsi_RO*, *proporsi_AC*, dan *proporsi_AE*.

2.8.2 Transformasi Logaritmik Waktu

Selain pola jawaban, penelitian ini juga memanfaatkan informasi waktu pengerjaan sebagai fitur tambahan yang menggambarkan perilaku pengguna selama menyelesaikan instrumen. Waktu pengerjaan dihitung berdasarkan durasi yang dibutuhkan pengguna dalam menjawab soal pada masing-masing dimensi gaya belajar.

Data waktu pengerjaan memiliki karakteristik distribusi yang cenderung tidak simetris dan dapat memiliki nilai ekstrem akibat perbedaan kecepatan pengerjaan antar pengguna. Oleh karena itu, dilakukan transformasi logaritmik untuk mengurangi pengaruh nilai ekstrem serta menghasilkan distribusi data yang lebih stabil [47].

Transformasi logaritmik dilakukan menggunakan persamaan:

$$\text{log\_waktu}_i = \ln(\text{waktu}_i + 1) \tag{2.11}$$

dengan:

- log_waktu_i adalah hasil transformasi waktu pada dimensi ke-i,
- waktu_i adalah durasi pengerjaan dalam satuan detik.

Melalui proses tersebut dibentuk lima fitur waktu, yaitu *log_waktu_CE*, *log_waktu_RO*, *log_waktu_AC*, *log_waktu_AE*, dan *log_waktu_total*.

2.8.3 Indeks Efisiensi

Selain menggunakan proporsi jawaban dan waktu pengerjaan secara terpisah, penelitian ini membentuk fitur indeks efisiensi yang menggabungkan kedua informasi tersebut. Indeks efisiensi digunakan untuk menggambarkan hubungan antara tingkat kecenderungan pengguna terhadap suatu dimensi gaya belajar dengan waktu yang dibutuhkan dalam memberikan respons.

Nilai efisiensi dihitung menggunakan perbandingan antara nilai proporsi jawaban dan log waktu pengerjaan pada dimensi yang sama.

$$\text{efisiensi}_i = \frac{\text{proporsi}_i}{\text{log\_waktu}_i} \tag{2.12}$$

Nilai efisiensi yang lebih tinggi menunjukkan bahwa pengguna memiliki kecenderungan yang lebih besar terhadap suatu dimensi gaya belajar dengan waktu pengerjaan yang relatif lebih singkat. Dengan demikian, fitur efisiensi dapat memberikan representasi tambahan mengenai karakteristik perilaku pengguna yang tidak hanya mempertimbangkan pola jawaban, tetapi juga aspek waktu pengerjaan.

Berdasarkan perhitungan tersebut, dibentuk empat fitur efisiensi, yaitu *efisiensi_CE*, *efisiensi_RO*, *efisiensi_AC*, dan *efisiensi_AE*.

2.8.4 Standardisasi Fitur

Setiap fitur yang digunakan dalam penelitian ini memiliki rentang nilai yang berbeda, sehingga dilakukan proses standardisasi sebelum digunakan dalam proses klasifikasi. Standardisasi dilakukan menggunakan metode *Z-score Standardization* [43] dengan persamaan:

$$z = \frac{x - \mu}{\sigma} \tag{2.13}$$

dengan:

- x adalah nilai asli suatu fitur,
- μ adalah nilai rata-rata fitur pada data pelatihan,

- σ adalah nilai simpangan baku fitur pada data pelatihan.

Proses standardisasi dilakukan setelah pembagian dataset menjadi data pelatihan dan data pengujian. Nilai rata-rata dan simpangan baku dihitung hanya berdasarkan data pelatihan, kemudian parameter tersebut digunakan untuk mentransformasikan data pengujian. Pendekatan ini dilakukan untuk mencegah terjadinya *data leakage* selama proses pembelajaran model.

Standardisasi diperlukan karena fitur yang digunakan memiliki rentang nilai yang berbeda. Fitur proporsi memiliki rentang nilai antara 0 hingga 1, sedangkan fitur log waktu pengerjaan dan efisiensi dapat memiliki rentang nilai yang lebih besar. Perbedaan skala tersebut dapat memengaruhi proses estimasi distribusi probabilitas pada algoritma Gaussian Naive Bayes. Melalui transformasi Z-*score*, seluruh fitur memiliki skala yang lebih seragam dengan rata-rata mendekati nol dan simpangan baku mendekati satu sehingga proses pembelajaran model menjadi lebih stabil [43, 47].

2.9 Perilaku Pengerjaan Soal sebagai Indikator Karakteristik Belajar

Di lingkungan pembelajaran secara digital, data tentang cara siswa berinteraksi bisa digunakan untuk mengerti bagaimana cara belajar mereka. Selain keakuratan jawaban, durasi waktu dalam menyelesaikan soal juga menjadi indikator penting yang mencerminkan proses berpikir yang terjadi saat belajar. Beberapa penelitian di bidang *Educational Data Mining* menunjukkan bahwa waktu yang dibutuhkan seseorang untuk mengerjakan soal bisa digunakan untuk mengetahui cara belajarnya, tingkat pemahamannya, serta ciri-ciri pribadi dalam menyelesaikan tugas belajar [37, 38]. Selain itu, siswa yang bisa mencapai hasil yang baik dalam waktu yang tidak terlalu lama dianggap lebih efisien dalam belajar dibandingkan siswa yang butuh waktu lebih lama untuk mencapai hasil yang sama [40].