BAB 2
Tinjauan Pustaka

2.1 Penelitian Terdahulu

Tinjauan terhadap penelitian terdahulu dilakukan untuk memetakan perkembangan mutakhir dalam bidang pemantauan pola tidur berbasis *machine learning*, mengidentifikasi celah (*gap*) yang belum terjawab oleh studi-studi sebelumnya, serta memosisikan kontribusi penelitian ini secara akademis dalam lanskap literatur ilmiah terkini. Lima penelitian yang paling relevan diulas secara kritis pada bagian berikut.

Rahman et al. [6] mengembangkan sistem klasifikasi tahap tidur (*sleep stage classification*) menggunakan sinyal *electroencephalography* (EEG) satu kanal dengan algoritma *Random Forest*. Penelitian tersebut memberikan kontribusi penting dalam membuktikan bahwa *Random Forest* mampu mengklasifikasikan tahap tidur dengan akurasi tinggi hanya dari satu kanal sensor, sehingga mereduksi kompleksitas perangkat dibandingkan sistem *polysomnography* (PSG) konvensional [7]. Kendati demikian, pendekatan ini mengandung keterbatasan mendasar yang membatasi penerapannya secara luas di masyarakat. Sistem tersebut tetap bergantung secara mutlak pada perangkat sensor EEG yang memerlukan pemasangan elektroda khusus di kepala, sehingga pengguna tidak dapat memanfaatkannya secara mandiri tanpa bantuan tenaga medis atau fasilitas klinis. Di samping itu, penelitian tersebut hanya berfokus pada klasifikasi tahap tidur sesaat tanpa menganalisis defisit tidur yang terakumulasi dari waktu ke waktu, serta sama sekali tidak menghasilkan keluaran berupa rekomendasi kesehatan yang dapat ditindaklanjuti secara langsung oleh pengguna.

Sathyanarayana et al. [8] mengusulkan metode prediksi kualitas tidur dengan memanfaatkan data dari perangkat *wearable* menggunakan pendekatan *deep learning*. Penelitian ini memberikan kontribusi signifikan dalam mendemonstrasikan bahwa data akselerometer dari perangkat komersial portabel dapat dieksploitasi untuk memprediksi pola kualitas tidur secara otomatis. Namun demikian, terdapat dua kendala utama yang membatasi jangkauan adopsi dari metode ini. Pertama, sistem mewajibkan kepemilikan perangkat keras pendukung berupa gelang atau jam tangan cerdas yang harganya relatif mahal, sehingga tidak dapat diakses secara merata oleh seluruh kalangan masyarakat. Kedua, model

arsitektur jaringan saraf tiruan yang digunakan bersifat sebagai kotak hitam (*black-box*), di mana proses pengambilan keputusan internalnya sulit diinterpretasikan secara medis [9]. Akibatnya, hasil prediksi tersebut menjadi kurang memadai apabila harus dijadikan landasan argumen rekomendasi kesehatan yang transparan dan dapat dipahami oleh orang awam.

Zhang et al. [10] menerapkan jaringan saraf konvolusional ortogonal (*Orthogonal Convolutional Neural Networks*) untuk otomatisasi klasifikasi tahap tidur berbasis sinyal EEG satu kanal dan berhasil mencapai tingkat akurasi sebesar 0,91. Inovasi utama dari penelitian tersebut terletak pada penerapan prinsip orthogonalitas pada lapisan konvolusional untuk menekan redundansi fitur serta menjaga stabilitas proses pelatihan model. Meskipun performa teknis yang dicapai sangat tinggi, keterbatasan mendasar dari sistem ini masih serupa dengan studi Rahman et al., yakni ketergantungan yang tinggi pada perangkat sensor EEG serta ketiadaan modul sistem rekomendasi. Lebih lanjut, interpretabilitas dari model konvolusional mendalam ini relatif rendah apabila dikomparasikan dengan algoritma berbasis pohon seperti *Random Forest*, karena representasi internal pada fitur tersembunyinya tidak memiliki pemetaan langsung ke variabel fisiologis yang bermakna secara klinis [5].

Bjorvatn et al. [11] melakukan penelitian mendalam mengenai dampak akumulasi utang tidur atau defisit tidur (*sleep debt*) terhadap peningkatan risiko infeksi melalui pendekatan kohort longitudinal selama satu tahun. Temuan penting dari penelitian ini mengungkapkan bahwa akumulasi defisit tidur yang melampaui ambang batas tertentu khususnya lebih dari dua jam berkontribusi secara signifikan terhadap kerentanan sistem imun tubuh. Studi ini memberikan fondasi teoretis dan klinis yang sangat kuat bagi penelitian yang sedang dikembangkan, di mana parameter kuantitatif tersebut diadopsi secara rasional untuk merumuskan batasan ambang batas tingkat keparahan *(severity level)* defisit tidur pada sistem, seperti kategori ringan hingga parah.

Fox et al. [12] memperluas cakupan analisis mengenai akumulasi utang tidur dengan menitikberatkan pada tinjauan demografis dan sosial. Mereka membuktikan bahwa variabel seperti kategori usia, jenis kelamin, dan latar belakang aktivitas harian memiliki korelasi langsung terhadap pola penumpukan defisit tidur seseorang. Temuan ini memperkuat urgensi bahwa sistem pemantauan tidur modern tidak boleh bersifat seragam *(size fits all)*, melainkan harus mampu memberikan rekomendasi personal yang sensitif terhadap karakteristik serta profil demografis masing-masing pengguna.

Berdasarkan tinjauan kritis terhadap kelima penelitian terdahulu tersebut, dapat disimpulkan bahwa penelitian ini berhasil mengisi celah ilmiah yang ada. Sistem yang dibangun memposisikan diri sebagai solusi inovatif yang beroperasi sepenuhnya secara mandiri melalui input manual berbasis web tanpa keharusan menggunakan sensor keras eksternal yang mahal. Dengan mengintegrasikan algoritma *Random Forest* yang memiliki interpretabilitas tinggi [5], sistem mampu mengolah selisih durasi tidur ideal dan aktual ke dalam pembagian kelas terstruktur (*normal*, *mild*, *moderate*, dan *severe*) serta menyajikan rekomendasi kesehatan yang dipersonalisasi secara akurat berdasarkan profil usia pengguna.

2.1.1 Analisis Kelemahan Penelitian Sebelumnya

Berdasarkan tinjauan kritis terhadap berbagai literatur mutakhir dalam domain pemantauan kualitas tidur dan kecerdasan buatan, teridentifikasi sejumlah celah serta kelemahan mendasar yang membatasi efektivitas studi-studi terdahulu. Analisis komprehensif mengenai kelemahan tersebut dijabarkan melalui poin-poin berikut:

- **Ketergantungan Tinggi pada Perangkat Keras dan Aksesibilitas Terbatas:** Sebagian besar penelitian terdahulu [6, 8, 10] sangat bergantung pada perangkat akuisisi data khusus, seperti elektroda *electroencephalography* (EEG) maupun gelang atau jam tangan cerdas (*wearable devices*). Ketergantungan ini menciptakan hambatan ekonomi dan teknis yang signifikan bagi masyarakat umum, sehingga sistem tidak dapat diakses secara praktis di lingkungan non-klinis atau digunakan secara mandiri tanpa intervensi tenaga medis.
- **Orientasi Model yang Terbatas pada Klasifikasi Sesaat Tanpa Analisis Defisit Kumulatif:** Studi-studi sebelumnya umumnya berfokus secara sempit pada klasifikasi tahap tidur instan (sleep stage classification) pada satu titik waktu tertentu. Pendekatan tersebut mengabaikan aspek temporal yang krusial, yaitu akumulasi utang tidur atau defisit tidur (*sleep debt*) yang menumpuk lintas waktu (seperti rentang 24, 48, hingga 72 jam), padahal dampak klinis terhadap penurunan kognitif dan kerentanan imun bersifat kumulatif [2, 11].
- **Sifat *Black-Box* Model dan Ketiadaan Personalisasi Rekomendasi:** Penerapan arsitektur komputasi kompleks seperti jaringan saraf tiruan

mendalam (*deep learning*) sering kali menghasilkan karakteristik kotak hitam (*black-box*) dengan tingkat interpretabilitas yang rendah [9]. Akibatnya, sistem gagal memberikan penjelasan logis yang transparan kepada pengguna. Selain itu, keluaran yang dihasilkan umumnya bersifat seragam (*one-size-fits-all*) tanpa memperhitungkan variabel demografis penting seperti rentang usia dan profil aktivitas harian pengguna [12].

- **Absennya Evaluasi Kuantitatif Kepuasan Pengguna Berbasis Instrumen Standar:** Mayoritas riset di bidang pemantauan tidur berbasis teknologi hanya menitikberatkan pengujian pada metrik performa algoritma semata (seperti akurasi klasifikasi), sementara aspek validasi kegunaan perangkat lunak dari perspektif end-user sangat minim. Penelitian terdahulu jarang mengintegrasikan evaluasi pengalaman pengguna yang terstruktur secara formal menggunakan kerangka pengukuran kuantitatif yang mapan, seperti metode *End-User Computing Satisfaction* (EUCS), guna menilai fungsionalitas dan kenyamanan antarmuka secara objektif.

2.1.2 Perbandingan Penelitian dan Gap

Tabel 2.1 merangkum perbandingan penelitian terdahulu dan posisi penelitian ini.

Tabel 2.1. Perbandingan penelitian terdahulu dan posisi penelitian ini

Aspek	**Rahman et al. (2022)**	**Sathyanarayana et al. (2016)**	**Zhang et al. (2019)**	**Penelitian Ini**
Algoritma	*Random Forest*	*Deep Learning*	CNN Ortogonal	*Random Forest*
Sumber data	EEG	*Wearable*	EEG	Input manual berbasis web
Platform	Klinis / desktop	Mobile / wearable	Klinis / desktop	Web browser
Analisis multi-periode	Tidak	Tidak	Tidak	Ya: 24/48/72 jam
Rekomendasi personal	Tidak	Tidak	Tidak	Ya, berbasis usia dan kualitas tidur
Interpretabilitas model	Sedang	Rendah	Rendah	Tinggi, dengan *feature importance*
Aksesibilitas tanpa sensor	Tidak	Tidak	Tidak	Ya

2.2 Teori Tidur dan Defisit Tidur

Bagian ini menjelaskan konsep tidur, kebutuhan tidur, dan kerangka severity level sebagai dasar klasifikasi dalam penelitian.

2.2.1 Fisiologi Tidur dan Fungsi Tidur

Tidur merupakan suatu proses fisiologis yang bersifat universal, kompleks, serta esensial bagi kelangsungan hidup organisme multiseluler. Secara biologis, tidur bukan sekadar keadaan pasif berupa ketiadaan kesadaran, melainkan sebuah siklus aktif yang diatur secara ketat oleh sistem saraf pusat untuk memfasilitasi proses pemulihan fisik dan pemeliharaan kognitif tingkat tinggi [13]. Selama siklus tidur berlangsung, tubuh mengalami serangkaian reorganisasi metabolik dan neuroendokrin yang krusial bagi regenerasi jaringan sel, penguatan sistem kekebalan tubuh, serta stabilitas emosional [14, 15].

Secara arsitektural, siklus tidur manusia umumnya terbagi menjadi dua fase utama yang saling bergantian secara periodik, yaitu fase *Non-Rapid Eye Movement* (NREM) dan fase *Rapid Eye Movement* (REM). Fase NREM yang terbagi menjadi beberapa tahapan mendalam memegang peranan utama dalam pemulihan fisik, perbaikan jaringan otot, serta pelepasan hormon pertumbuhan. Sementara itu, fase REM sangat diasosiasikan dengan aktivitas otak yang tinggi, pemrosesan memori emosional, serta konsolidasi memori jangka panjang [16].

Ketika individu mengalami gangguan tidur kronis atau akumulasi defisit tidur yang berkepanjangan, homeostasis tubuh akan mengalami disrupsi signifikan. Berbagai kajian epidemiologis dan klinis menunjukkan bahwa kekurangan tidur kronis berkorelasi kuat dengan penurunan fungsi eksekutif kognitif, penurunan waktu reaksi, serta gangguan regulasi metabolik yang memicu peningkatan risiko patologi serius, seperti hipertensi, penyakit kardiovaskular, obesitas, gangguan resistensi insulin, serta penurunan respons imun tubuh terhadap infeksi patogen [17, 18, 11].

2.2.2 Kebutuhan Tidur Berdasarkan Kelompok Usia

Kebutuhan tidur ideal berbeda berdasarkan usia. Tabel 2.2 menunjukkan kategori usia yang digunakan dalam sistem dan acuan durasi tidur yang relevan.

Tabel 2.2. Kebutuhan tidur ideal menurut kategori usia

Kategori Usia	**Rentang Usia**	**Kebutuhan Optimal (jam/hari)**
Balita	0–4 tahun	11,0
Anak-anak	5–11 tahun	10,0
Remaja	12–17 tahun	9,0
Dewasa	18 tahun ke atas	8,0

2.2.3 Defisit Tidur

Defisit tidur didefinisikan sebagai selisih antara durasi tidur ideal dan durasi tidur aktual (dalam satuan jam). Besaran nilai defisit tersebut dirumuskan sebagai berikut:

$$\text{Defisit Tidur} = \text{Durasi Tidur Ideal} - \text{Durasi Tidur Aktual} \tag{2.1}$$

Dengan Durasi Tidur Ideal merupakan standar jam tidur yang direkomendasikan berdasarkan kategori usia pengguna, dan Durasi Tidur Aktual adalah total jam tidur yang berhasil dicatat oleh pengguna dalam periode analisis.

Nilai selisih defisit tidur tersebut selanjutnya digunakan untuk menentukan tingkat keparahan *(severity level)* ke dalam empat kelas terstruktur sebagai berikut. Dalam implementasi sistem ini, aturan tersebut digunakan sebagai dasar pelabelan dataset selama preprocessing data training, sedangkan hasil akhir prediksi severity ditentukan oleh model Random Forest.

$$\text{Severity} = \begin{cases} \textit{Normal} & \text{jika Defisit} \leq 0 \text{ jam (Tidur cukup)} \\ \textit{Mild} & \text{jika Defisit} < 2 \text{ jam (Kurang ringan)} \\ \textit{Moderate} & \text{jika Defisit} \leq 3 \text{ jam (Kurang sedang)} \\ \textit{Severe} & \text{jika Defisit} > 3 \text{ jam (Kurang parah)} \end{cases} \tag{2.2}$$

Penggunaan ambang batas kuantitatif ini mendukung analisis defisit kumulatif lintas periode (seperti 24 jam, 48 jam, dan 72 jam). Dalam implementasi proyek ini, aturan ambang tersebut digunakan terutama untuk menghasilkan label training selama preprocessing data, sedangkan prediksi akhir severity pada runtime ditentukan oleh model *Random Forest* melalui mekanisme voting ensemble. Dengan demikian, sistem mampu mengidentifikasi bukan hanya kekurangan tidur pada satu malam saja, tetapi juga tren akumulasi utang tidur *(sleep debt)* yang menumpuk dari waktu ke waktu [2, 11].

Rujukan: Analisis defisit kumulatif dan pengaruhnya terhadap penurunan performa kognitif serta kesehatan fisiologis telah dikaji secara ekstensif dalam berbagai literatur ilmiah [19, 20].

2.2.4 Dampak Kesehatan Jangka Pendek dan Panjang

Insufisiensi tidur jangka pendek memengaruhi konsentrasi, reaksi, dan mood, sementara ketidakcukupan tidur jangka panjang meningkatkan risiko masalah metabolik dan kardiovaskular [2].

Penelitian ini memetakan kategori severity ke dalam dampak kesehatan yang disuport oleh knowledge base JSON, sehingga rekomendasi tidak hanya bersifat statistik, tetapi juga medis.

2.3 Teori Machine Learning dan Random Forest

Aplikasi *machine learning* dalam penelitian ini berfokus pada klasifikasi severity level berdasarkan pola tidur pengguna.

2.3.1 Decision Tree

Algoritma *Decision Tree* (Pohon Keputusan) adalah salah satu metode pembelajaran mesin terarah (supervised learning) yang bekerja dengan cara memecah kumpulan data yang kompleks menjadi struktur hierarkis yang lebih sederhana berupa pohon keputusan. Struktur ini terdiri dari tiga elemen utama: simpul akar (root nodee), simpul internal (internal node) yang merepresentasikan kondisi pengujian pada suatu fitur, serta daun (leaf node) yang menyatakan label kelas keputusan akhir [21].

Dalam proses pembangunannya, algoritma menggunakan ukuran pemilihan fitur seperti Information Gain, Gain Ratio, atau Gini Impurity untuk menentukan fitur mana yang paling optimal sebagai titik belah (split) pada setiap iterasi. Keunggulan utama dari *Decision Tree* terletak pada interpretabilitasnya yang sangat tinggi, di mana alur logika keputusan dari data mentah menuju hasil akhir dapat divisualisasikan dan dipahami secara intuitif oleh manusia tanpa memerlukan pemahaman matematis yang rumit [22]. Kendati demikian, model tunggal (single tree) rentan mengalami overfitting apabila pohon tumbuh terlalu dalam, sehingga sensitif terhadap perubahan kecil pada data latih.

2.3.2 Ensemble Learning: Random Forest

Untuk mengatasi kelemahan model tunggal yang rentan terhadap overfitting serta memiliki variansi yang tinggi, Leo Breiman mengembangkan metode Ensemble Learning yang dikenal sebagai *Random Forest* [5]. *Random Forest* bekerja dengan cara menggabungkan prediksi dari banyak pohon keputusan (ensemble of decision trees) yang dibangun secara paralel dan independen.

Pendekatan ini mengandalkan dua teknik utama dalam proses pelatihannya:

1. ***Bootstrap Aggregating* (Bagging):** Setiap pohon keputusan dalam model dilatih menggunakan sampel data subset hasil pengambilan acak dengan pengembalian dari keseluruhan data latih, sehingga setiap pohon menerima variasi distribusi data yang unik.

2. **Pemilihan Fitur Acak (*Random Feature Selection*):** Pada setiap titik pemisahan (split) dalam pembuatan pohon, algoritma hanya memilih acak sebagian kecil dari total fitur yang tersedia (biasanya bernilai akar kuadrat dari total fitur), yang bertujuan untuk mengurangi korelasi antar pohon individu.

Melalui mekanisme penggabungan suara mayoritas (majority voting) untuk tugas klasifikasi, *Random Forest* mampu meredam variansi secara signifikan, meningkatkan ketahanan terhadap *noise*, serta menghasilkan performa generalisasi yang jauh lebih unggul dan stabil dibandingkan model Decision Tree tunggal [23].

2.3.3 Bootstrap Sampling

Metode *Bootstrap Sampling* adalah teknik resampling statistik yang digunakan untuk menghasilkan beberapa subset data latih dari satu set data utama dengan cara pengambilan sampel secara acak dengan pengembalian (sampling with replacement). Melalui teknik ini, sebuah sampel memiliki peluang yang sama untuk dipilih kembali atau bahkan tidak dipilih sama sekali.Secara teoretis, probabilitas sebuah sampel data tunggal tidak terpilih sama sekali dalam satu kali proses pembuatan sampel bootstrap (dengan ukuran n sampel) didefinisikan melalui persamaan limit berikut:

$$P(\text{tidak terpilih}) = \left(1 - \frac{1}{n}\right)^n \approx e^{-1} \approx 0{,}368 \tag{2.3}$$

Persamaan di atas menunjukkan bahwa secara rata-rata, sekitar 36,8% dari total data asli akan menjadi sampel out-of-bag (OOB) dan tidak digunakan dalam pelatihan pohon keputusan tertentu, sehingga dapat dimanfaatkan secara optimal untuk proses validasi internal model tanpa memerlukan set validasi terpisah [24].

2.3.4 Pemilihan Fitur Acak

Dalam arsitektur *Random Forest*, proses pengacakan tidak hanya dilakukan pada level data melalui bootstrap sampling, tetapi juga pada level fitur (random feature selection). Pada setiap simpul (split) dalam pohon keputusan, algoritma tidak mengevaluasi seluruh fitur yang tersedia, melainkan hanya memilih sebagian kecil fitur secara acak. Pendekatan ini memastikan bahwa setiap pohon individu memiliki keragaman struktural yang tinggi, serta mampu meningkatkan ketahanan

model (robustness) terhadap multikolinearitas dan fitur-fitur yang saling berkorelasi erat.

2.3.5 Gini Impurity

Metrik *Gini Impurity* adalah ukuran statistik yang digunakan untuk mengevaluasi tingkat ketidakmurnian atau ketidakteraturan (impurity) dari suatu himpunan data pada simpul pohon keputusan. Nilai ini merepresentasikan probabilitas kesalahan klasifikasi dari elemen yang dipilih secara acak jika diberi label sesuai dengan distribusi kelas dalam himpunan tersebut. Gini Impurity dirumuskan sebagai berikut:

$$Gini(S) = 1 - \sum_{k=1}^{K} p_k^2 \tag{2.4}$$

Dengan K menyatakan jumlah total kelas dan p_k adalah proporsi sampel yang termasuk dalam kelas k pada himpunan data S. Nilai Gini bernilai 0 jika seluruh sampel dalam simpul tersebut murni berasal dari satu kelas yang sama.

2.3.6 Weighted Gini Split

Untuk menentukan titik pemisahan (split point) yang paling optimal pada suatu fitur, algoritma menghitung tingkat pengurangan ketidakmurnian melalui nilai rata-rata berbobot dari impurity simpul anak sebelah kiri dan kanan. Nilai Weighted Gini Split dirumuskan sebagai berikut:

$$Gini_{split} = \frac{|S_L|}{|S|} Gini(S_L) + \frac{|S_R|}{|S|} Gini(S_R) \tag{2.5}$$

Di mana $|S|$ adalah jumlah total sampel pada simpul induk, serta $|S_L|$ dan $|S_R|$ masing-masing menyatakan jumlah sampel pada simpul anak kiri dan kanan setelah proses pemisahan.

2.3.7 Majority Voting

Sebagai sebuah model ensemble yang terdiri dari banyak pohon keputusan independen, proses pengambilan keputusan akhir untuk tugas klasifikasi pada *Random Forest* dilakukan melalui mekanisme pemungutan suara mayoritas

(majority voting). Keputusan akhir ditentukan oleh kelas yang paling sering diprediksi oleh keseluruhan pohon, yang secara matematis dituliskan sebagai:

$$\hat{y} = \arg\max_{c \in C} \sum_{t=1}^{T} \mathbf{1}(h_t(x) = c) \tag{2.6}$$

Dengan T adalah total jumlah pohon dalam forest, C adalah himpunan kelas target, $h_t(x)$ adalah prediksi kelas oleh pohon ke-t, dan $\mathbf{1}(\cdot)$ merupakan fungsi indikator yang bernilai 1 jika kondisi di dalamnya benar dan 0 jika sebaliknya.

2.3.8 Interpretabilitas dan Penjelasan Model

Meskipun model ensemble sering kali dianggap lebih rumit dibandingkan model tunggal, *Random Forest* tetap menyediakan mekanisme analitis berupa *feature importance* untuk mengukur kontribusi relatif masing-masing variabel masukan terhadap hasil prediksi global. Dalam implementasi praktis sistem ini, hanya *feature importance* yang digunakan sebagai ukuran interpretabilitas utama, dengan memetakan kontribusi fitur secara langsung ke dalam rekomendasi kesehatan yang dapat dipahami pengguna.

Pendekatan interpretabilitas modern seperti *Local Interpretable Model-agnostic Explanations* (LIME) dan *Shapley Additive exPlanations* (SHAP) dikenal luas sebagai metode pelengkap dalam sistem cerdas, tetapi belum diintegrasikan dalam implementasi ini dan diposisikan sebagai potensi pengembangan selanjutnya [25, 26, 27].

2.3.9 Evaluasi Klasifikasi

Performa generalisasi dari model klasifikasi yang dibangun diukur menggunakan empat metrik evaluasi utama, yaitu *Accuracy, Precision, Recall*, dan *F1-Score*. Untuk skenario klasifikasi biner maupun analisis per-kelas, metrik-metrik tersebut didefinisikan melalui persamaan matematis berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.7}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.8}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.9}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.10}$$

Di mana TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative) merujuk pada elemen-elemen yang terkandung di dalam matriks konfusi (confusion matrix) [9, 28].

2.3.10 Macro Average

Dalam evaluasi multi-kelas, penggunaan *Macro Average (Macro-Averaging)* dilakukan untuk menghitung nilai rata-rata aritmetika dari metrik evaluasi (seperti *precision, recall*, atau *F1-score*) di seluruh kelas tanpa memandang ketidakseimbangan jumlah sampel per kelas. Hal ini memastikan bahwa setiap kelas, baik kelas mayoritas maupun minoritas, mendapatkanbobot kontribusi yang setara dalam penilaian performa global model.

2.3.11 Confusion Matrix dan Metrik Per-Kelas

Evaluasi komprehensif terhadap performa klasifikasi tingkat keparahan defisit tidur disajikan menggunakan *confusion matrix* untuk merinci ketepatan prediksi antar kelas. Contoh bentuk matriks konfusi untuk empat kategori kelas (*normal*, *mild*, *moderate*, dan *severe*) diilustrasikan pada Tabel 2.3.

2.3.12 Confusion Matrix dan Metrik Per-Kelas

Evaluasi komprehensif terhadap performa klasifikasi tingkat keparahan defisit tidur disajikan menggunakan *confusion matrix* untuk merinci ketepatan prediksi antar kelas. Contoh bentuk matriks konfusi untuk empat kategori kelas (*normal*, *mild*, *moderate*, dan *severe*) diilustrasikan pada Tabel 2.3.

Tabel 2.3. Contoh Confusion Matrix untuk Klasifikasi Severity

	Pred: Normal	**Pred: Mild**	**Pred: Moderate**	**Pred: Severe**
Actual: Normal	50	3	1	0
Actual: Mild	4	30	2	0
Actual: Moderate	1	5	18	1
Actual: Severe	0	0	2	12

Berdasarkan penjabaran matriks konfusi tersebut, kalkulasi metrik performa per-kelas dievaluasi secara terperinci seperti yang dirangkum pada Tabel 2.4.

Tabel 2.4. Contoh Metrik Per-Kelas (Precision / Recall / F1)

Kelas	Precision	Recall	F1-Score
Normal	0.91	0.94	0.92
Mild	0.75	0.83	0.79
Moderate	0.78	0.72	0.75
Severe	0.92	0.86	0.89

2.4 Sistem Rekomendasi dalam Domain Kesehatan

Sistem rekomendasi berbasis komputer dalam ranah pelayanan kesehatan dan kebugaran memegang peranan krusial dalam menjembatani kesenjangan antara literatur klinis yang kompleks dengan tindakan preventif sehari-hari yang dapat dilakukan oleh masyarakat awam. Agar efektif, sebuah sistem rekomendasi cerdas dituntut untuk mampu menyeimbangkan tingkat personalisasi yang tinggi dengan landasan keandalan ilmiah yang dapat dipertanggungjawabkan secara medis [29].

2.4.1 Rule-Based dan Hybrid Recommender

Dalam arsitektur sistem pendukung keputusan kesehatan, pendekatan konvensional umumnya mengandalkan sistem berbasis aturan *(rule-based recommender systems)* yang mengevaluasi kondisi pengguna secara deterministik berdasarkan sekumpulan logika *if-then* yang disusun oleh pakar domain. Meskipun pendekatan ini sangat transparan dan minim kesalahan logika fatal, sistem berbasis aturan murni sering kali kaku dan kurang adaptif terhadap pola data pengguna yang kompleks [30].

Sebaliknya, pendekatan hibrida *(hybrid recommender systems)* menggabungkan keunggulan logika aturan eksplisit dengan kemampuan generalisasi dari model pembelajaran mesin *(machine learning)* serta penyimpanan basis pengetahuan *(knowledge base)* yang terstruktur [31]. Dalam kerangka penelitian ini, sistem memanfaatkan rekomendasi dasar yang bersumber dari repositori basis pengetahuan berbasis format JSON, yang kemudian dipadukan secara dinamis dengan hasil klasifikasi tingkat keparahan (severity) dari model *Random Forest* serta penyesuaian personal berdasarkan profil aktual pengguna.

2.4.2 Personalisasi Berdasarkan Profil Pengguna

Personalisasi merupakan elemen inti yang menentukan relevansi dan kemanjuran suatu intervensi kesehatan digital. Pendekatan *one-size-fits-all* atau generalisasi perlakuan kesehatan sering kali gagal karena kebutuhan fisiologis individu sangat bervariasi bergantung pada tahapan usia, ritme sirkadian, serta gaya hidup.

Dalam sistem yang dikembangkan, personalisasi tidak hanya berfokus pada durasi tidur malam tunggal, melainkan memperhitungkan kategori rentang usia, tingkat kualitas tidur, serta konsistensi akumulasi pola defisit lintas periode. Pendekatan ini selaras secara empiris dengan temuan Fox et al. [12], yang menekankan bahwa variabel demografis dan sosial memiliki korelasi yang signifikan terhadap bagaimana utang tidur *(sleep debt)* terakumulasi dan memengaruhi performa fisiologis seseorang dalam populasi modern.

2.4.3 Transparansi Rekomendasi

Dalam domain yang sensitif seperti kesehatan, aspek transparansi dan interpretabilitas *(explainability)* dari sebuah rekomendasi merupakan syarat mutlak. Keluaran sistem tidak boleh bersifat sebagai kotak hitam (black-box) yang memberikan instruksi tanpa alasan logis, karena hal tersebut dapat membahayakan keselamatan pengguna dan menurunkan tingkat kepercayaan terhadap teknologi [32].

Dengan memanfaatkan algoritma *Random Forest* yang dilengkapi dengan analisis tingkat kepentingan fitur (feature importance) serta struktur basis pengetahuan JSON yang terpetakan dengan jelas, sistem mampu menautkan setiap saran tindakan medis preventif kembali ke faktor-faktor penentu yang bermakna secara klinis. Hal ini memastikan bahwa setiap rekomendasi perubahan gaya hidup atau waktu istirahat yang diterima oleh pengguna memiliki justifikasi ilmiah yang transparan, logis, dan mudah dipahami.

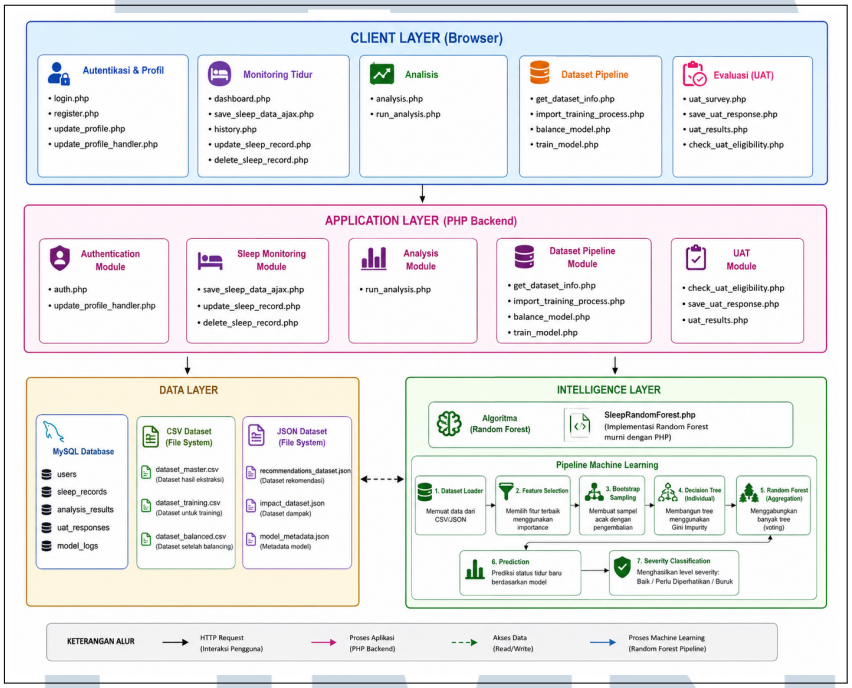
2.5 Aplikasi Web dan Basis Data

Aplikasi web ini dirancang agar dapat diakses melalui peramban dan memproses data tidur secara terstruktur. Periode 24 jam, 48 jam, dan 72 jam dihitung berdasarkan agregasi catatan tidur harian yang diinput manual oleh

pengguna pada tabel `sleep_records`, bukan melalui pengukuran langsung dari perangkat sensor.

2.5.1 Arsitektur Aplikasi

Arsitektur aplikasi terdiri dari frontend, backend PHP, database MySQL, dan knowledge base JSON. Diagram arsitektur dapat ditempatkan pada Gambar 2.1.



Gambar 2.1. Diagram arsitektur sistem aplikasi pemantauan pola tidur berbasis web.

2.5.2 Modul Utama Implementasi

Modul aplikasi terdiri dari:

- Autentikasi dan profil pengguna: `login.php`, `register.php`, `update_profile.php`, `update_profile_handler.php`.
- Input dan histori tidur: `dashboard.php`, `save_sleep_data_ajax.php`, `history.php`, `update_sleep_record.php`, `delete_sleep_record.php`.

- Analisis data dan model: `analysis.php`, `run_analysis.php`, `algorithms/SleepRandomForest.php`.

- Pipeline dataset training: `get_dataset_info.php`, `train_model.php`, `balance_model.php`, `import_training_process.php`.

- Evaluasi pengguna: `uat_survey.php`, `save_uat_response.php`, `uat_results.php`, `check_uat_eligibility.php`.

2.5.3 Desain Basis Data

Basis data MySQL menyimpan tabel utama: `users`, `sleep_records`, `analysis_results`, `sleep_training_data`, dan `uat_responses`. Desain ini mendukung pelacakan profil, histori tidur, hasil analisis, dan evaluasi.

Catatan: Skema basis data juga mendefinisikan tabel seperti `health_impacts` dan `medical_recommendations` untuk potensi penyimpanan referensi medis terstruktur. Namun, implementasi runtime saat ini menggunakan file `datasets/impact_dataset.json` dan `datasets/recommendations_dataset.json` sebagai sumber rekomendasi dan dampak.

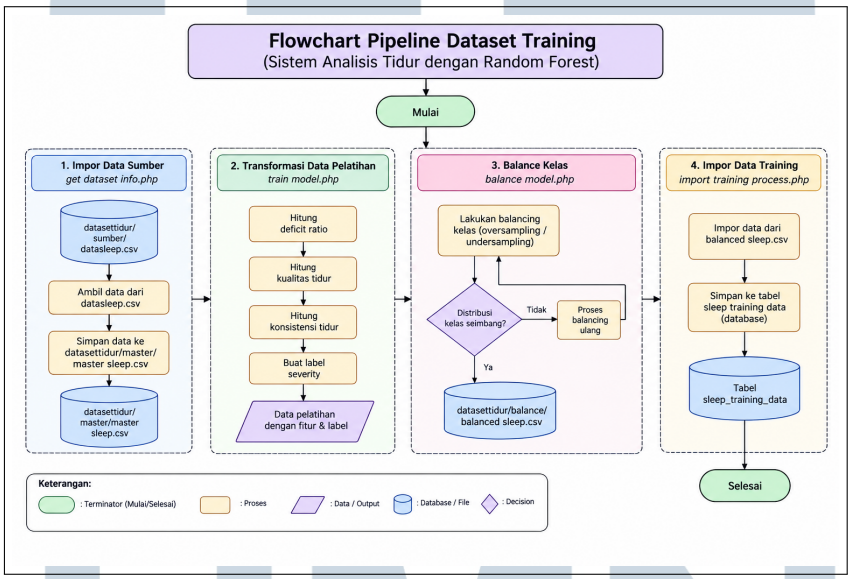
2.5.4 Pipeline Data dan Dataset

Proses data berjalan melalui beberapa tahap:

1. Impor data sumber dari `datasleep.csv` yang berisi sekitar 15.000 sampel ke `master_sleep.csv` melalui `get_dataset_info.php`.
2. Transformasi data pelatihan di `train_model.php` untuk menghitung *deficit_ratio*, kualitas, konsistensi, dan label severity.
3. Balance kelas melalui `balance_model.php` untuk menghasilkan `datasettidur/balance/balanced_sleep.csv` dengan distribusi seimbang, menghasilkan total akhir 6.000 sampel (1.500 data per kategori).

4. Data periode 48 jam dan 72 jam dibentuk secara derivatif dari data 24 jam dengan penyesuaian durasi dan kolom `period`, bukan dari catatan input pengguna selama dua atau tiga hari berturut-turut. Pendekatan ini adalah batasan metodologis yang perlu dipahami dalam konteks pengolahan data training.

5. Impor data training ke tabel `sleep_training_data` menggunakan `import_training_process.php`.

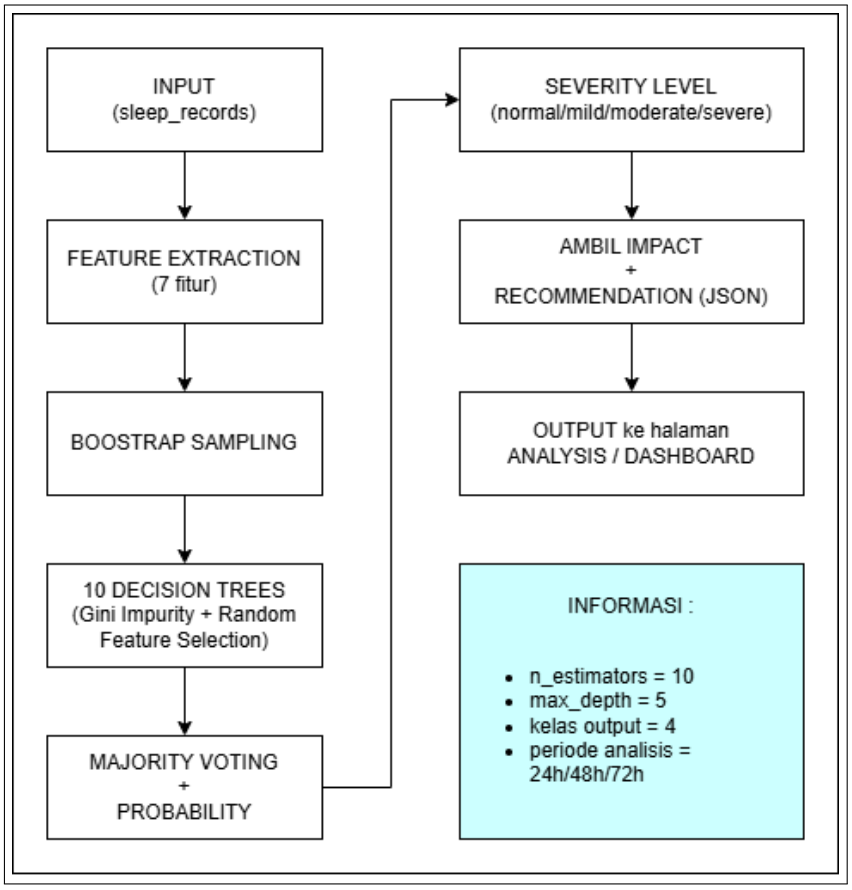


Gambar 2.2. Alur pemrosesan data training dan inferensi model Random Forest.

2.5.5 Komponen Algoritma

Komponen algoritma dalam implementasi terdiri dari:

- kelas `DecisionTreeNode`,
- kelas `DecisionTree`,
- kelas `SleepRandomForest`,
- fungsi *feature extraction*,
- fungsi prediksi kelas, probabilitas, dan *feature importance*,
- fungsi evaluasi model *hold-out* dengan *confusion matrix*.



Gambar 2.3. Alur teknis pemrosesan *Random Forest* pada sistem.

2.5.6 Komponen Visualisasi dan Interaksi

Visualisasi menggunakan *Chart.js* untuk:

- grafik tren tidur 7 hari di *dashboard*,
- grafik histori 30 hari di halaman *history*,
- visualisasi *feature importance* pada halaman *analysis*.

Interaksi data menggunakan mekanisme *AJAX*/*fetch* sehingga pembaruan analisis dapat ditampilkan cepat tanpa perpindahan halaman.

2.5.7 Interaksi Pengguna dan AJAX

Frontend mengirim data melalui AJAX ke backend dalam format JSON atau form data. Hal ini dipakai untuk menyimpan input tidur, menjalankan analisis, dan memeriksa kelayakan UAT tanpa memuat ulang halaman.

2.6 Keterkaitan Teori dengan Implementasi Proyek

Tabel 2.5 menunjukkan keterkaitan antara konsep theoretis dan implementasi kode yang aktual.

Tabel 2.5. Keterkaitan konsep teoritis dengan implementasi kode

Konsep	**Implementasi**	**Berkas Kode**
Input tidur	Input tanggal, jam tidur, jam bangun, kualitas, dan perhitungan durasi otomatis	`dashboard.php`, `save_sleep_data_ajax.php`
Histori dan manajemen data	Tabel histori, edit/hapus record, grafik tren tidur	`history.php`, `update_sleep_record.php`, `delete_sleep_record.php`
Analisis multi-periode	Ekstraksi fitur, prediksi 24/48/72 jam, confidence	`run_analysis.php`, `SleepRandomForest.php`
Rekomendasi	Knowledge base JSON dan personalisasi berdasarkan kualitas/konsistensi	`recommendations_dataset.json`, `impact_dataset.json`
Evaluasi pengguna	Instrumen EUCS, penyimpanan jawaban, dan skor dimensi	`uat_survey.php`, `save_uat_response.php`, `uat_results.php`

2.7 Novelty dan Kontribusi Penelitian

Penelitian ini menangani beberapa gap utama dalam literatur:

- Aksesibilitas tanpa sensor khusus melalui aplikasi web.
- Deteksi defisit tidur kumulatif pada tiga periode berbeda.
- Integrasi model klasifikasi dengan rekomendasi kesehatan yang dipersonalisasi.
- Evaluasi kepuasan pengguna menggunakan kerangka EUCS secara komprehensif.

Tabel 2.6. Perbandingan fitur utama penelitian ini dengan studi terdahulu

Aspek	**Sensor bebas**	**Analisis 24/48/72h**	**Rekomendasi personal**	**EUCS**
Penelitian EEG / wearable	Tidak	Tidak	Tidak	Tidak
Sistem web umum	Ya	Tidak	Tergantung	Tidak
Penelitian ini	Ya	Ya	Ya	Ya

2.8 Ringkasan Gap yang Ditangani

Penelitian ini mengintegrasikan elemen yang biasanya terpisah dalam literatur menjadi satu sistem operasional, yaitu:

- Aksesibilitas melalui web tanpa perangkat tambahan,
- Analisis defisit tidur kumulatif 24/48/72 jam,
- Rekomendasi personal berdasarkan profil pengguna,
- Evaluasi kepuasan pengguna menggunakan EUCS.

2.9 *End-User Computing Satisfaction* (EUCS)

Kepuasan pengguna akhir dievaluasi menggunakan kerangka *End-User Computing Satisfaction* (EUCS) yang dikembangkan oleh Doll dan Torkzadeh [33]. EUCS mengukur kepuasan pengguna terhadap sistem informasi berbasis komputer melalui enam dimensi:

1. *Content* (Isi Informasi): relevansi dan kelengkapan informasi yang disajikan sistem.
2. *Accuracy* (Keakuratan): ketepatan dan konsistensi data serta hasil analisis yang dihasilkan sistem.
3. *Format* (Bentuk Tampilan): kualitas estetika dan keterbacaan antarmuka pengguna.
4. *Ease of Use* (Kemudahan Penggunaan): tingkat kemudahan dalam mengoperasikan seluruh fitur sistem.

5. *Timeliness* (Ketepatan Waktu): kecepatan respons sistem dalam menyajikan informasi dan hasil analisis.

6. *User Satisfaction* (Kepuasan Pengguna): penilaian kepuasan menyeluruh terhadap sistem.

Instrumen EUCS pada penelitian ini terdiri dari 30 butir pernyataan (indeks A1–Y5) dengan skala Likert 1–5, di mana indeks A1 berfungsi sebagai filter kualitatif dan indeks A2–Y5 digunakan sebagai data kuantitatif [34]. Skor per dimensi dihitung sebagai rata-rata aritmetika indeks dalam kelompok yang sama, sedangkan skor keseluruhan dihitung dari rata-rata seluruh indeks Likert (A2–Y5).

Pengolahan hasil EUCS pada implementasi proyek tersusun dalam modul server: berkas save_uat_response.php melakukan validasi jawaban (rentang 1 sampai 5) dan menyimpan hasil ke dalam tabel uat_responses, sementara berkas uat_results.php menyediakan ringkasan skor per dimensi dan tabel distribusi. Struktur ini direpresentasikan pada Tabel 2.7 yang juga ditampilkan pada antarmuka hasil evaluasi.

Tabel 2.7. Indeks evaluasi EUCS dan interpretasi skor

Dimensi EUCS	**Rentang Skor**	**Interpretasi**
Content (Isi Informasi)	1.0–2.5	Kurang relevan / tidak lengkap
Accuracy (Keakuratan)	2.6–3.5	Cukup akurat, perlu validasi tambahan
Format (Tampilan)	3.6–4.2	Tampilan baik, mudah dibaca
Ease of Use (Kemudahan)	3.0–4.0	Mayoritas pengguna menemukan sistem mudah digunakan
Timeliness (Ketepatan Waktu)	2.5–4.5	Respons sistem memadai untuk sebagian besar pengguna
User Satisfaction (Kepuasan)	2.8–4.5	Tingkat kepuasan moderat hingga tinggi

Contoh pemetaan skor keseluruhan ke kategori interpretasi:

- 1.0–2.0: Kurang (perlu perbaikan signifikan)

- 2.1–3.4: Cukup (perlu peningkatan pada beberapa aspek)
- 3.5–4.2: Baik (layak direkomendasikan)
- 4.3–5.0: Sangat Baik (rekomendasi kuat)

Hasilnya, penelitian ini tidak hanya memperluas teori, tetapi juga menyajikan implementasi praktis dan transparan yang dapat diuji lebih lanjut.