

BAB 2 LANDASAN TEORI

Pada bab ini akan dijabarkan penjelasan teoritis sebagai dasar dalam penelitian ini. Teori-teori tersebut meliputi konsep tidur dan insomnia, pengantar *machine learning*, serta penjelasan detail mengenai algoritma *Decision Tree* dan *Logistic Regression*.

2.1 Penelitian Pendahulu

Dalam penelitian ini peneliti melakukan penelaahan terhadap beberapa studi terdahulu yang memiliki kesamaan terkait metode atau topik dengan penelitian yang dilakukan. Tabel berikut merupakan hasil tabulasi penelitian terdahulu yang menjadi acuan dalam penyusunan penelitian.

Tabel 2.1. Tabel Penelitian Pendahulu

No	Penulis	Judul	Algoritma	Hasil
1	Khansa et al. (2024)[11]	Analisis Perbandingan Algoritma <i>Machine Learning</i> Dalam Klasifikasi Gangguan Tidur	<i>Naïve Bayes</i> , <i>Support Vector Machine</i> (SVM), dan <i>Neural Network</i>	Berdasarkan evaluasi, algoritma <i>Neural Network</i> memiliki performa tertinggi akurasi sebesar 93,08%.
2	Diana Sari. (2024)[12]	Prediksi Gangguan Tidur pada <i>Sleep Health and Lifestyle</i> Menggunakan <i>Support Vector Machine</i> dan <i>Neural Network</i>	<i>Support Vector Machine</i> dan <i>Neural Network</i>	Hasil akurasi tertinggi berasal dari metode <i>Neural Network</i> sebesar 91.2%

Tabel 2.1 Tabel Penelitian Pendahulu (lanjutan)

No	Penulis	Judul	Algoritma	Hasil
3	Maulidah et al. (2024)[13]	PREDIKSI KESEHATAN TIDUR DAN GAYA HIDUP MENGGUNAKAN MACHINE LEARNING	Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Gradient Boosting, dan Decision Tree	Hasil penelitian menunjukkan bahwa <i>Gradient Boosting</i> memberikan akurasi tertinggi sebesar 91%
4	Galuh Dwi Rahayuningtyas (2025)[14]	Prediksi <i>Sleep Disorder</i> Menggunakan Metode Algoritma <i>Naïve Bayes Gaussian</i>	<i>Naïve Bayes Gaussian</i>	Berdasarkan pengujian didapatkan akurasi sebesar 85,3%
5	Faradillah et al. (2025)[15]	Model Prediksi Gangguan Tidur berdasarkan Beberapa Faktor menggunakan <i>Machine Learning</i>	<i>Logistic Regression, Decision Tree</i> dan <i>Gradient Boosting</i>	Model <i>Gradient Boosting</i> menghasilkan akurasi prediksi tertinggi, dengan akurasi 99%

Penelitian oleh Khansa et al. bertujuan untuk memanfaatkan *dataset Sleep Health and Lifestyle* dari platform *Kaggle* untuk membangun model klasifikasi gangguan tidur seperti *insomnia* dan *sleep apnea*. Dalam penelitian ini, tiga algoritma *data mining* dibandingkan kinerjanya, yaitu *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Neural Network*. Metodologi yang digunakan mengacu pada

tahapan *CRISP-DM*, dimulai dari pemahaman data, persiapan data, hingga evaluasi model. Proses evaluasi dilakukan menggunakan teknik *cross validation*, *confusion matrix*, dan kurva ROC. Hasil penelitian ini menunjukkan bahwa algoritma *Neural Network* memberikan performa terbaik dalam mengklasifikasi gangguan tidur, dengan akurasi mencapai 93,08% dan nilai AUC yang masuk dalam kategori “*Excellent*”. Temuan penelitian terdahulu ini menegaskan bahwa *Neural Network* memiliki potensi tinggi dalam mendukung deteksi dini gangguan tidur secara otomatis, sehingga mampu membantu proses diagnosis klinis dengan lebih akurat dan efisien.

Penelitian yang dilakukan oleh Diana Sari menyoroti pentingnya faktor gaya hidup dalam memengaruhi kualitas tidur dan munculnya gangguan tidur seperti insomnia. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang berkontribusi terhadap gangguan tidur seseorang, seperti beban pekerjaan, durasi tidur, kualitas tidur, aktivitas fisik, dan kebiasaan hidup sehari-hari. Penelitian ini juga memanfaatkan pendekatan *data mining* dalam membangun model prediksi berbasis dua algoritma klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Neural Network*, yang kemudian diimplementasikan menggunakan *platform Orange*. Hasil eksperimen menunjukkan bahwa algoritma *Neural Network* unggul dengan akurasi 91,2%, dibandingkan SVM yang memperoleh akurasi sebesar 90,1%. Temuan ini membuktikan bahwa *Neural Network* merupakan metode yang efektif untuk klasifikasi data kesehatan tidur dan gaya hidup serta dapat dijadikan referensi dalam pengembangan sistem prediksi gangguan tidur secara akurat dan efisien.

Penelitian Maulidah et al. menghasilkan bahwa *dataset Sleep Health and Lifestyle* dari *Kaggle* digunakan untuk menganalisis hubungan gaya hidup dengan gangguan tidur menggunakan berbagai algoritma *machine learning*. Dari beberapa model yang diuji, *Gradient Boosting* menunjukkan akurasi tertinggi sebesar 91%, diikuti oleh *Random Forest* dan SVM dengan akurasi 88%. Faktor seperti aktivitas fisik, tingkat stres, dan BMI ditemukan berpengaruh signifikan terhadap kualitas tidur.

Penelitian oleh Galuh Dwi Rahayuningtyas membahas prediksi gangguan tidur seperti insomnia dan *sleep apnea* menggunakan algoritma klasifikasi *Naïve Bayes Gaussian*. Gangguan ini, jika tidak ditangani, dapat menimbulkan komplikasi serius seperti hipertensi dan depresi. Melalui pemodelan prediktif, penelitian ini memperoleh akurasi sebesar 85,3%, yang menunjukkan bahwa algoritma *Naïve Bayes* cukup efektif dalam mengidentifikasi gangguan tidur secara dini, sehingga dapat membantu proses penanganan secara cepat dan tepat.

Penelitian yang dilakukan oleh Faradillah et al. mengembangkan model prediksi gangguan tidur berbasis *machine learning* dengan memanfaatkan data sekunder dari *Kaggle*. Beberapa algoritma seperti *logistic regression*, *decision tree*, dan *gradient boosting* diterapkan, dengan hasil menunjukkan bahwa *Gradient Boosting* memberikan akurasi tertinggi sebesar 99%.

Model *machine learning* akan dicoba dalam 2 algoritma yaitu *Decision Tree* dan *Logistic Regression*, hal ini berlandaskan pada penelitian sebelumnya pada November 2024 telah dilaksanakannya perbandingan untuk algoritma *Naïve Bayes*, *Support Vector Machine* dan *Neural Network* dengan hasil yang cukup mengesankan, dengan algoritma *Neural Network* memiliki tingkat akurasi 93,08% dibandingkan 2 algoritma lainnya [11].

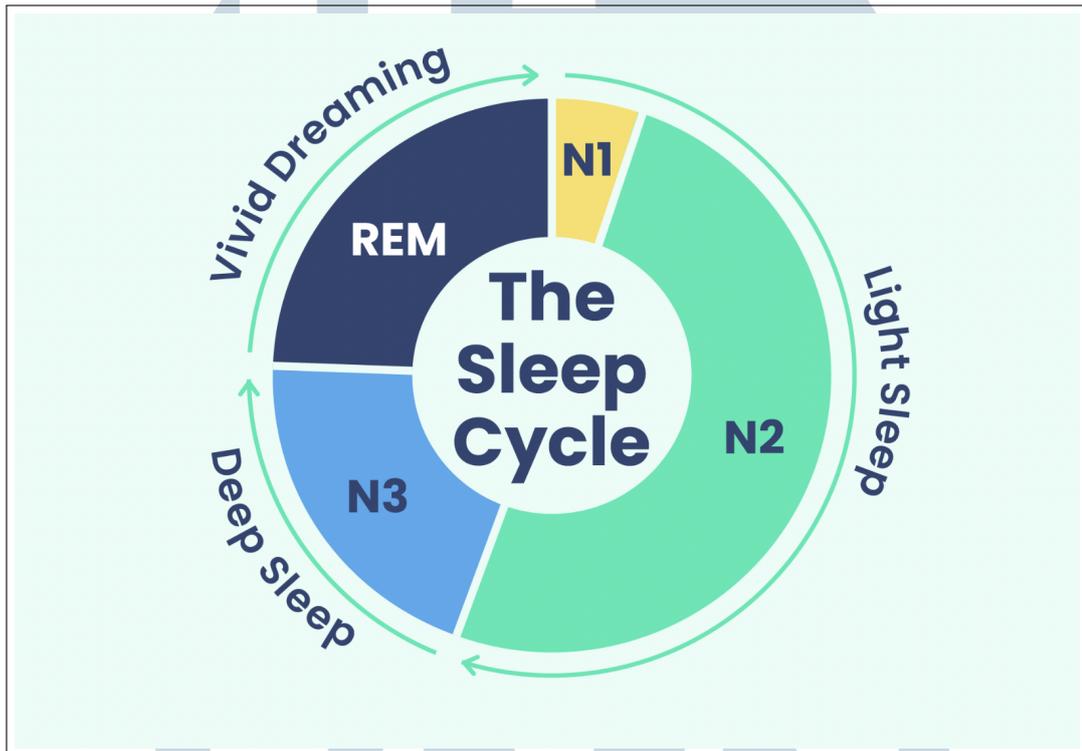
Melalui beberapa penelitian yang telah di telaah oleh peneliti, dikumpulkan beberapa kesenjangan penelitian (*research gap*) yang menjadi dasar penting dalam penelitian ini. Pertama, sebagian besar studi sebelumnya dalam prediksi insomnia menggunakan data klinis atau survei langsung, sedangkan penggunaan data dari perangkat *wearable* yang diperluas dengan teknik augmentasi seperti *Vector Autoregression (VAR)* masih sangat jarang diterapkan [16, 17]. Pendekatan ini membuka peluang baru dalam meningkatkan kuantitas dan variasi data kesehatan, terutama dalam konteks gangguan tidur. Kedua, meskipun algoritma kompleks seperti *Neural Network* dan *Support Vector Machine* telah banyak diteliti, studi yang membandingkan secara langsung performa dua algoritma yang lebih sederhana namun *interpretable* — yaitu *Decision Tree* dan *Logistic Regression* — dalam prediksi insomnia masih terbatas [18]. Padahal, kedua algoritma ini memiliki keunggulan dari sisi transparansi hasil dan efisiensi, yang sangat relevan untuk implementasi di bidang kesehatan. Ketiga, masih minimnya penelitian yang mengombinasikan data kebiasaan tidur dari sumber terbuka dengan pendekatan sintetik untuk membangun model prediksi yang akurat dan dapat digunakan secara luas juga menjadi celah yang ingin dijawab oleh studi ini.

2.2 Tidur

Tidur merupakan aktivitas fisiologis yang sangat penting bagi keberlangsungan hidup manusia. Proses tidur yang optimal berperan dalam memulihkan energi tubuh, memperkuat sistem imun, serta memperbaiki fungsi otak, termasuk dalam hal konsolidasi memori dan pengendalian emosi [19].

Secara biologis, tidur dibagi menjadi dua fase besar: *Non-Rapid Eye*

Movement (NREM) dan *Rapid Eye Movement (REM)*, yang berlangsung secara siklikal selama tidur. *NREM* terbagi ke dalam tiga tahapan, yaitu tidur ringan hingga tidur dalam, yang memiliki fungsi penting dalam pemulihan fisik. Sementara itu, fase *REM* berperan dalam pemrosesan informasi dan memori emosional [20]. Untuk lebih jelasnya dapat dilihat pada gambar 2.1:



Gambar 2.1. Siklus Tidur Manusia
Sumber: *Sleep Foundation*, 2025 [21]

Gambar 2.1 di atas menggambarkan siklus tidur manusia yang terdiri dari empat fase utama: N1, N2, N3, dan *REM*. N1 menandai awal tidur ringan, diikuti oleh N2 yang lebih dalam dan menjadi fase terpanjang. N3 adalah tidur dalam yang penting untuk pemulihan fisik. *REM* ditandai dengan mimpi intens dan aktivitas otak yang tinggi. Siklus ini berulang beberapa kali setiap malam, dan kualitas tidur ditentukan oleh keseimbangan antar fase tersebut.

Tabel 2.2. Siklus Tidur dalam Bentuk Tabel

Sumber: *Sleep Foundation*, 2025 [21]

Tahap Tidur	Jenis Tidur	Panjang Rata-rata
Tahap 1	<i>NREM</i>	1 - 7 menit
Tahap 2	<i>NREM</i>	10 - 25 menit
Tahap 3	<i>NREM</i>	20 - 40 menit
Tahap 4	<i>REM</i>	10 - 60 menit

Tabel 2.2 di atas menjelaskan empat tahap tidur yang diklasifikasikan berdasarkan jenis dan durasi rata-ratanya. Tahap 1 hingga 3 termasuk dalam tidur *NREM*, dengan durasi berturut-turut antara 1–7 menit, 10–25 menit, dan 20–40 menit. Tidur *NREM* berperan penting dalam proses pemulihan fisik dan transisi dari sadar ke tidur dalam.

Sementara itu, tahap 4 merupakan tidur *REM* yang berlangsung sekitar 10–60 menit, ditandai dengan aktivitas mimpi dan peningkatan aktivitas otak. Setiap tahap memiliki peran fisiologis berbeda dan berlangsung secara berulang dalam siklus tidur malam hari.

Tidur yang tidak berkualitas dapat menyebabkan gangguan kognitif, penurunan imunitas, dan meningkatkan risiko penyakit kronis seperti hipertensi, obesitas, dan diabetes [22]. Rata-rata waktu tidur ideal bagi orang dewasa adalah 7 hingga 9 jam per malam, dan kurang tidur kronis dapat berdampak pada produktivitas, kestabilan emosi, dan kesehatan mental seseorang [23].

2.3 Kebiasaan Tidur

Kebiasaan tidur dapat mencakup rutinitas harian sebelum tidur, durasi dan waktu tidur, kualitas lingkungan tidur, serta aktivitas yang memengaruhi kondisi tubuh menjelang tidur. Faktor-faktor seperti konsumsi kafein, paparan cahaya biru dari gawai, stres psikologis, dan gaya hidup yang tidak aktif dapat memperburuk kualitas tidur seseorang [24].

Kebiasaan tidur yang buruk seperti tidur larut malam secara konsisten, menggunakan perangkat elektronik menjelang tidur, atau tidur dalam kondisi lingkungan yang bising dan terang, terlihat dapat menurunkan kualitas dan durasi tidur. Berikut penjelasan lebih dalam terkait faktor-faktor yang mempengaruhi

kebiasaan tidur:

1. Denyut Jantung

Denyut jantung menjadi indikator utama dalam menentukan kondisi fisiologis tubuh, termasuk dalam fase tidur. Variabilitas denyut jantung yang tinggi biasanya mencerminkan tubuh dalam keadaan rileks dan siap untuk tidur. Sebaliknya, denyut jantung yang cepat dan tidak teratur dapat mengindikasikan stres atau gangguan pada sistem saraf otonom, sehingga mengganggu *onset* dan kualitas tidur [25].

2. Temperatur Badan

Suhu inti tubuh manusia secara alami menurun saat mendekati waktu tidur dan mencapai titik terendah di tengah malam [26]. Penurunan suhu ini merupakan sinyal biologis yang penting bagi otak untuk memulai dan mempertahankan tidur. Lingkungan tidur yang terlalu panas atau terlalu dingin dapat mengganggu proses termoregulasi tubuh dan menghambat penurunan suhu inti yang diperlukan untuk tidur nyenyak [27]. Kamar yang terlalu panas, misalnya, dapat menyebabkan tubuh terus berusaha mendinginkan diri, mengganggu arsitektur tidur dan menyebabkan terbangun di malam hari [28].

3. Gerakan Saat Tidur

Meskipun beberapa gerakan tubuh selama tidur adalah normal (misalnya, berbalik posisi), gerakan yang berlebihan atau tidak terkontrol dapat menjadi indikator gangguan tidur. Gerakan ini bisa berupa *restless legs syndrome* (sindrom kaki gelisah) yang menyebabkan dorongan tak tertahankan untuk menggerakkan kaki, atau *periodic limb movement disorder* (gangguan gerakan anggota tubuh periodik) yang melibatkan gerakan kaki atau lengan secara berulang saat tidur [29]. Gerakan-gerakan ini dapat mengganggu kelangsungan tidur, menyebabkan sering terbangun, dan menurunkan kualitas tidur secara keseluruhan.

4. Konsumsi Kafein

Kafein adalah stimulan sistem saraf pusat yang bekerja dengan memblokir *adenosin*, zat kimia otak yang memicu rasa kantuk [30]. Ketika *adenosin* diblokir, kita merasa lebih terjaga dan energik. Namun, kafein memiliki waktu paruh yang relatif panjang (sekitar 3-7 jam) [31], yang berarti efeknya bisa bertahan lama di dalam tubuh. Mengonsumsi kafein terlalu dekat dengan

waktu tidur dapat menunda inisiasi tidur, mengurangi total waktu tidur, dan mengurangi kualitas tidur. Sensitivitas terhadap kafein juga bervariasi antar individu [32].

5. **Tingkat Stres**

Stres mengaktifkan respons "lawan atau lari" dalam tubuh, melepaskan hormon seperti *kortisol* dan adrenalin. Hormon-hormon ini meningkatkan kewaspadaan dan membuat tubuh siap menghadapi ancaman, yang sangat berlawanan dengan kondisi yang dibutuhkan untuk tidur [33]. Pikiran yang berpacu, kekhawatiran, dan ketegangan fisik akibat stres dapat membuat sulit untuk tertidur dan sering terbangun di malam hari [34]. Stres kronis dapat mengganggu siklus tidur-bangun alami tubuh dan menyebabkan gangguan tidur jangka panjang.

6. **Perubahan Jam Tidur**

Tubuh manusia memiliki jam internal atau ritme *sirkadian* yang mengatur siklus tidur-bangun. Ritme ini sangat dipengaruhi oleh paparan cahaya dan kegelapan, serta kebiasaan tidur yang konsisten. Perubahan jam tidur yang drastis atau tidak teratur (misalnya, begadang hingga pagi) dapat mengganggu ritme *sirkadian* ini. Akibatnya, tubuh akan kesulitan untuk mengetahui kapan harus tidur dan kapan harus bangun, menyebabkan insomnia, kelelahan di siang hari, dan masalah kesehatan lainnya [35]. Konsistensi dalam jadwal tidur adalah kunci untuk menjaga ritme *sirkadian* yang sehat.

7. **Paparan Cahaya**

Cahaya, terutama cahaya biru yang dipancarkan oleh perangkat elektronik (ponsel, tablet, laptop), memiliki pengaruh kuat terhadap ritme *sirkadian* [35]. Paparan cahaya di malam hari, terutama sebelum tidur, dapat menekan produksi *melatonin*, hormon yang memicu rasa kantuk dan mengatur siklus tidur. Ketika produksi *melatonin* terganggu, tubuh akan kesulitan untuk tertidur dan tidur dapat menjadi kurang berkualitas [36].

- Studi mahasiswa UPI Bandung (2024) menemukan 38,6% mengalami gangguan tidur saat lampu menyala, dan 81,4% kualitas tidur membaik saat gelap [37].
- Studi dari Semarang UIA (2022) menunjukkan cahaya lampu sebelum tidur menghambat *melatonin* dan ritme *sirkadian*, menyebabkan kualitas tidur buruk [38].

8. Durasi Tidur

Tidur dengan durasi yang tidak normal, baik terlalu singkat maupun terlalu lama, berhubungan dengan berbagai masalah kesehatan. Oleh karena itu, durasi tidur menjadi salah satu komponen terpenting dari kesehatan tidur secara keseluruhan [39]. Menurut penelitian, durasi tidur yang baik untuk manusia sekitar 7 - 9 jam. Sebaliknya, jika durasi tidur kurang dari 7 jam, maka tidur dianggap kurang berkualitas/ insomnia [12].

2.4 Insomnia

Insomnia adalah gangguan tidur yang ditandai dengan kesulitan untuk memulai tidur, mempertahankan tidur, atau bangun terlalu dini dan tidak dapat kembali tidur. Alhasil gejala dari gangguan tidur ini bisa menjadi gangguan fungsional saat bangun dan beraktivitas pada siang hari [40].

Secara klinis, insomnia dikategorikan sebagai gangguan tidur kronis apabila berlangsung setidaknya tiga kali dalam seminggu selama periode tiga bulan berturut-turut. Definisi ini merujuk pada kriteria diagnostik dalam *International Classification of Sleep Disorders, Third Edition (ICSD-3)* dari *American Academy of Sleep Medicine* [41].

2.4.1 Prevalensi Insomnia

Berdasarkan laporan dari *National Sleep Foundation* (2018), sekitar 67% populasi global mengalami gejala insomnia, termasuk di antaranya 1.508 individu yang tercatat berasal dari kawasan Asia Tenggara. Di Indonesia sendiri, angka prevalensi insomnia diperkirakan mencapai 28% dari total populasi sebanyak 238 juta jiwa, atau setara dengan sekitar 10% dari seluruh masyarakat secara nasional[7]. Berarti terdapat sekitar 28 juta individu mengalami masalah insomnia[42]. Di antara penderita insomnia tersebut:

- Sekitar 51,8% mengalami insomnia ringan, ditandai dengan keluhan sesekali namun berdampak minimal terhadap fungsi harian [43].
- Sekitar 25,9% mengalami insomnia sedang, yang sudah mengganggu fungsi sosial, akademik, dan psikologis secara nyata [43].

2.4.2 Gejala dan Tanda-Tanda Klinis

Secara umum, insomnia ditandai oleh berbagai gejala khas yang dapat dikenali sebagai indikasi gangguan tidur, di antaranya meliputi:

- Sulit untuk tertidur, terutama setelah waktu tidur yang diharapkan [44].
- Sering terbangun di malam hari dan membutuhkan waktu lama untuk kembali tidur [45].
- Bangun terlalu pagi dan tidak merasa segar [45].
- Kesulitan berkonsentrasi, gangguan memori, dan mudah marah [45].
- Merasa lelah dan mengantuk di siang hari [46].

Pada kasus kronis, insomnia juga berkaitan dengan peningkatan risiko gangguan psikologis seperti depresi, gangguan kecemasan, serta gangguan metabolik seperti hipertensi dan penyakit jantung.

2.4.3 Dampak Insomnia

Insomnia berdampak luas terhadap berbagai aspek kehidupan. Dari segi fisiologis hingga produktivitas tiap individu. Diantaranya dijelaskan melalui tabel 2.3:

Tabel 2.3. Dampak Insomnia Berdasarkan Aspeknya
Sumber: Jurnal Pengembangan Ilmu dan Praktik Kesehatan [47]

Dampak	Keterangan
Fisiologis	Gangguan keseimbangan hormonal, peningkatan tekanan darah, risiko obesitas
Psikologis	Meningkatkan risiko depresi, gangguan kecemasan, dan burnout
Kognitif	Menurunnya kemampuan fokus, memori kerja, dan kecepatan berpikir
Sosial dan Produktivitas	Absensi kerja, penurunan produktivitas, konflik <i>interpersonal</i>

Tabel 2.3 memperlihatkan beragam dampak insomnia yang terbagi dalam beberapa kategori. Dampak tersebut meliputi segi fisiologis yang dapat mengubah kondisi fisik seseorang, segi psikologis yang mengganggu kesehatan mental, segi kognitif yang menurunkan kemampuan fokus, serta segi sosial dan produktivitas yang berpotensi memicu konflik antarindividu.

2.5 Machine Learning

Machine Learning (ML) merupakan salah satu pendekatan dalam bidang kecerdasan buatan (*Artificial Intelligence*) yang dirancang untuk meniru kemampuan manusia dalam menganalisis data, belajar dari pola yang ada, serta menghasilkan keputusan atau prediksi secara otomatis. *Machine learning* memungkinkan sistem komputer untuk menyelesaikan permasalahan kompleks tanpa perlu diprogram secara eksplisit, melainkan melalui proses pembelajaran dari data yang tersedia. Konsep *machine learning* pertama kali diperkenalkan oleh Arthur Samuel pada tahun 1959 sebagai "kemampuan komputer untuk belajar tanpa diprogram secara eksplisit" [48].

Dalam penerapan *machine learning*, terdapat tiga kategori utama pendekatan pembelajaran, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. Ketiga pendekatan ini memiliki karakteristik serta keunggulan yang berbeda, sebagai berikut [49]:

1. *Supervised Learning*

Supervised learning adalah pendekatan pembelajaran yang mengandalkan data berlabel untuk melatih algoritma, sehingga sistem dapat belajar mengenali hubungan antara variabel *input* (X) dan target *output* (Y). Metode ini pada dasarnya menggunakan pendekatan fungsional, yaitu mencari fungsi terbaik yang dapat memetakan *input* ke *output* secara akurat.

Namun, dalam praktiknya, menemukan fungsi yang optimal seringkali tidak mudah karena prosesnya sangat dipengaruhi oleh asumsi dasar yang digunakan oleh algoritma. Jika asumsi tersebut tidak sesuai dengan karakteristik data, hasil prediksi bisa mengalami bias. Oleh karena itu, keberhasilan metode ini sangat bergantung pada kualitas dan ketepatan data latih yang digunakan.

Algoritma yang termasuk dalam *supervised learning* antara lain: regresi

linier, regresi logistik, *K-Nearest Neighbors (K-NN)*, *Naive Bayes*, *Decision Tree*, *Support Vector Machine (SVM)*, dan jaringan saraf tiruan (*artificial neural networks*). Kelebihan utama dari pendekatan ini adalah mudah dipahami dan efektif digunakan untuk tugas klasifikasi serta regresi. Namun, kekurangannya adalah memerlukan proses pelabelan data yang bisa memakan waktu dan membutuhkan sumber daya yang besar, serta proses pelatihan yang bisa berlangsung lama, terutama pada *dataset* berskala besar.

2. *Unsupervised Learning*

Berbeda dengan *supervised learning*, metode *unsupervised learning* tidak memerlukan data berlabel dalam proses pelatihannya. Algoritma ini bertujuan untuk menemukan pola atau struktur tersembunyi dalam data tanpa adanya target *output* tertentu. Oleh karena itu, pendekatan ini sering digunakan dalam tugas seperti *clustering* (pengelompokan data) dan *association rule mining* (pencarian pola asosiasi antar atribut).

Kelebihan dari *unsupervised learning* adalah fleksibilitasnya dalam mengeksplorasi pola yang belum diketahui sebelumnya tanpa perlu proses pelabelan yang mahal. Namun, kelemahannya terletak pada kesulitan dalam mengevaluasi hasil, karena tidak adanya label referensi sebagai pembanding. Hal ini membuat interpretasi hasil dari model *unsupervised* cenderung lebih kompleks.

3. *Semi-Supervised dan Reinforcement Learning*

Selain dua pendekatan utama di atas, terdapat pula metode *semi-supervised learning*, yaitu kombinasi antara *supervised* dan *unsupervised learning*. Metode ini cocok digunakan ketika hanya sebagian data yang berlabel, sementara sisanya tidak. *Semi-supervised learning* menawarkan solusi efisien dalam kondisi keterbatasan data berlabel, karena hanya sebagian kecil data yang memerlukan pelabelan manual oleh ahli. Pendekatan ini umumnya digunakan dalam skenario *big data* dengan anggaran terbatas untuk *labeling*. Sementara itu, *reinforcement learning* merupakan pendekatan pembelajaran berbasis *trial-and-error* yang meniru cara makhluk hidup belajar dari pengalaman. Dalam metode ini, terdapat agen (*agent*) yang berinteraksi dengan lingkungan (*environment*) untuk mengambil tindakan tertentu.

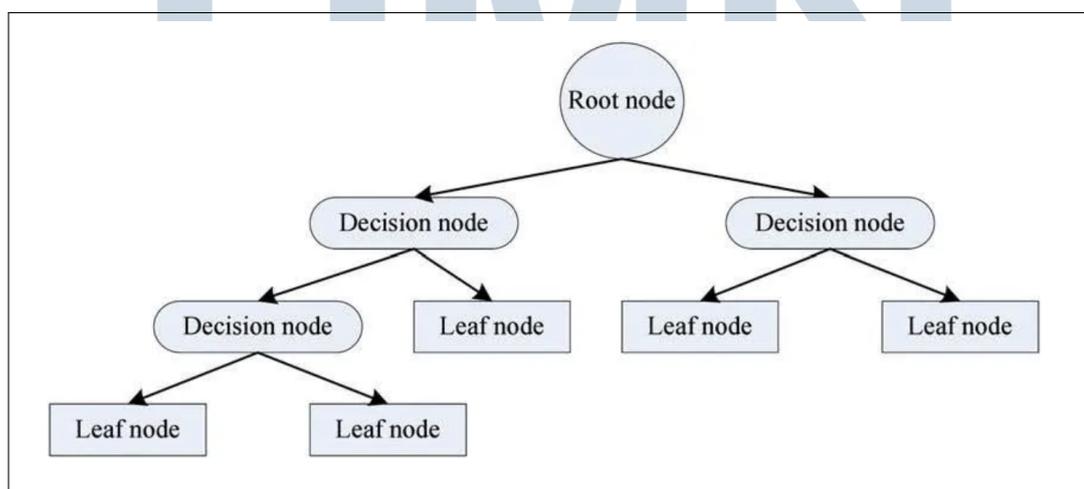
Berdasarkan tindakan yang diambil, agen akan menerima umpan balik berupa *reward* (penguatan positif) atau *punishment* (penguatan negatif). Proses ini berlangsung secara berulang-ulang, sehingga agen dapat mempelajari strategi optimal untuk memaksimalkan keuntungan dan meminimalkan risiko dalam jangka panjang.

Reinforcement learning sangat cocok diterapkan dalam sistem pengambilan keputusan dinamis, seperti robotika, permainan, dan sistem rekomendasi berbasis adaptasi lingkungan.

Penelitian ini menggunakan pendekatan *supervised learning* untuk klasifikasi biner (insomnia/tidak insomnia). Model yang akan digunakan adalah model *Decision Tree* dan *Logistic Regression*.

2.6 Decision Tree

Decision Tree atau pohon keputusan merupakan salah satu metode dalam pembelajaran mesin yang digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Algoritma ini bekerja dengan prinsip *divide and conquer*, yaitu membagi permasalahan kompleks menjadi bagian-bagian yang lebih sederhana dalam bentuk struktur pohon [50]. Setiap simpul (*node*) pada pohon merepresentasikan atribut atau fitur dari data, cabang (*branch*) menggambarkan keputusan berdasarkan nilai dari fitur tersebut, dan daun (*leaf node*) merupakan *output* akhir berupa kelas atau nilai prediksi [51].



Gambar 2.2. *Decision Tree Classifier*

Sumber: Medium [52]

Gambar 2.2 menggambarkan proses kerja algoritma *Decision Tree*, yang dimulai dari *root node* sebagai representasi keseluruhan data. Algoritma kemudian memilih fitur paling signifikan untuk membagi data dan membentuk pertanyaan berstruktur *if-else*, hingga menghasilkan prediksi pada *leaf node* [51].

Proses pembentukan pohon keputusan dilakukan dengan memilih fitur terbaik untuk membagi data berdasarkan ukuran tertentu, seperti *Gini Index*. Tujuannya adalah memperoleh pembagian yang menghasilkan klasifikasi paling informatif. Dengan pendekatan ini, *Decision Tree* sangat cocok digunakan untuk data yang memiliki hubungan non-linear, serta mampu menangani baik data numerik maupun kategorikal. Kemampuannya dalam menangani nilai kosong (*missing values*) juga menjadikan model ini cukup tangguh dan fleksibel dalam berbagai skenario analisis data [53].

Rumus 2.1 *Gini Index*.

$$GiniIndex = 1 - \sum_{i=1}^n (P_i)^2 = 1 - [(P_+)^2 + (P_-)^2] \quad (2.1)$$

Rumus 2.1 menunjukkan perhitungan *Gini Index*, di mana P_+ adalah probabilitas kelas positif dan P_- probabilitas kelas negatif. *Gini Index* dihitung untuk semua kemungkinan pemisahan, dan nilai terendah (*low impurity*) akan dipilih [54].

2.7 Vector Autoregression (VAR)

Vector Autoregression (VAR) merupakan salah satu model statistik multivariat yang digunakan untuk memodelkan dan menangkap hubungan linier antara beberapa variabel deret waktu (*time series*). Tidak seperti model regresi linier klasik yang memiliki satu variabel dependen dan beberapa variabel independen, VAR memperlakukan semua variabel sebagai endogen dan memodelkan setiap variabel sebagai fungsi linier dari lag-nya sendiri serta lag dari variabel lainnya [55].

Model VAR banyak digunakan dalam ekonomi, keuangan, dan sekarang juga di bidang teknologi kesehatan, terutama ketika data memiliki sifat temporal atau sekuensial. Dalam konteks penelitian ini, model VAR digunakan sebagai metode augmentasi data, yaitu untuk menghasilkan data sintetis berdasarkan pola dari data asli yang dikumpulkan dari perangkat *wearable*.

Secara umum, model VAR dengan p lag dinyatakan sebagai:

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \varepsilon_t \quad (2.2)$$

di mana:

- Y_t adalah vektor variabel endogen pada waktu t ,
- c adalah vektor konstanta (intersep),
- A_i adalah matriks koefisien untuk lag ke- i ,
- ε_t adalah vektor dari error terms (gangguan) yang diasumsikan white noise.

Melalui simulasi dari model yang telah dilatih, data sintetis dapat dihasilkan sehingga menambah jumlah data latih tanpa harus mengumpulkan data baru secara manual. Teknik ini bermanfaat dalam mengatasi keterbatasan jumlah data serta meningkatkan performa algoritma machine learning dalam prediksi risiko insomnia.

2.8 Standard Scaler

Standard Scaler adalah teknik normalisasi yang banyak dipakai. Cara kerjanya adalah dengan mencari nilai rata-rata dan standar deviasi data terlebih dahulu, lalu setiap nilai dalam data tersebut dibagi dengan standar deviasinya [?].

Proses *Standard Scaling* diterapkan pada seluruh fitur *input* (X), baik pada *data training*, validasi, maupun *testing*. Hal ini penting karena algoritma seperti *Logistic Regression* sensitif terhadap skala data. Rumus standardisasi yang digunakan adalah [56]:

Rumus 2.3 *Standard Scaler*.

$$z = \frac{x - \mu}{\sigma} \quad (2.3)$$

Keterangan:

- z : nilai fitur setelah dinormalisasi
- x : nilai asli dari fitur
- μ : nilai rata-rata dari fitur

- σ : standar deviasi dari fitur

Berikut adalah rumus untuk menghitung nilai rata-rata (*mean*), sedangkan penjelasan dari setiap komponennya dapat dilihat pada Rumus 2.4 [57].

Rumus 2.4 *Mean*.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.4)$$

Keterangan:

- μ : simbol yang digunakan untuk rata-rata dari *dataset*.
- n : jumlah total elemen atau observasi dalam *dataset*.
- x_i : Elemen-elemen data individu dalam *dataset*, dengan i sebagai indeks yang merepresentasikan urutan dari 1 hingga n .

Sementara itu, rumus untuk menghitung standar deviasi beserta penjelasan dari setiap komponennya dapat ditemukan pada Rumus 2.5 [57].

Rumus 2.5 Standar Deviasi.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (2.5)$$

Keterangan:

- σ : Merupakan notasi yang digunakan untuk menyatakan standar deviasi dari sebuah *dataset*.
- n : Menunjukkan total jumlah elemen atau observasi dalam *dataset*.
- x_i : Setiap elemen data individu, di mana i merupakan indeks yang berjalan dari 1 hingga n .
- μ : Menunjukkan nilai rata-rata dari seluruh elemen x_i dalam *dataset*.

2.9 Logistic Regression

Regresi logistik adalah metode analisis yang digunakan untuk melihat hubungan antara satu atau lebih variabel bebas dengan variabel terikat yang bersifat

dikotomis. Analisis ini memiliki penerapan luas di berbagai bidang, seperti kedokteran, pemasaran, dan ilmu sosial [58].

Regresi logistik menggunakan fungsi *sigmoid* untuk menghitung probabilitas kejadian, menghasilkan *output* biner 0 atau 1. Fungsi ini mengubah nilai kontinu menjadi rentang 0–1, sehingga dapat diinterpretasikan sebagai probabilitas. Metode ini cocok untuk memprediksi dua kelas berbeda [59].

Rumus 2.6 Fungsi *Sigmoid Logistic Regression*.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2.6)$$

Rumus 2.2 menunjukkan fungsi $\sigma(z)$ yang merupakan fungsi *sigmoid* yang mengubah *input* linear z menjadi nilai antara 0 hingga 1. Nilai z adalah masukan yang diterapkan ke dalam fungsi tersebut. Ketika z bernilai besar, maka e^{-z} menjadi sangat kecil sehingga $\sigma(z)$ mendekati 1, yang menunjukkan probabilitas tinggi terhadap variabel terikat. Sebaliknya, jika z bernilai kecil, maka e^{-z} akan besar sehingga $\sigma(z)$ mendekati 0, menandakan probabilitas yang rendah terhadap variabel terikat [60].

2.10 Evaluasi: Confusion Matrix

Confusion Matrix adalah metode evaluasi model klasifikasi yang membandingkan hasil prediksi dengan label sebenarnya. Metode ini terdiri dari empat komponen utama: *True Positive*, *True Negative*, *False Positive*, dan *False Negative*, yang digunakan untuk menghitung metrik evaluasi seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, untuk menilai kinerja model secara menyeluruh [61].

Tabel 2.4. Tabel *Confusion Matrix*

Sumber: Jurnal Ilmiah Komputer dan Informatika [62]

	Prediksi Positif	Prediksi Negatif
Kenyataan Positif	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Kenyataan Negatif	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Tabel 2.4 menunjukkan *confusion matrix*, yaitu matriks evaluasi yang digunakan untuk menilai performa model klasifikasi dalam membedakan dua kelas,

yaitu positif dan negatif. Matriks ini terdiri dari empat komponen utama: *True Positive (TP)*, yaitu jumlah data yang diprediksi positif dan benar-benar positif; *False Negative (FN)*, yaitu data yang sebenarnya positif namun diprediksi sebagai negatif; *False Positive (FP)*, yaitu data yang sebenarnya negatif namun diprediksi positif; dan *True Negative (TN)*, yaitu data yang diprediksi negatif dan memang negatif [61].

Keempat nilai ini sangat penting untuk mengukur kinerja model secara menyeluruh. Dari *confusion matrix*, dapat dihitung berbagai metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* [63]. Dengan demikian, I tidak hanya memberikan informasi tentang jumlah prediksi yang benar atau salah, tetapi juga menunjukkan jenis kesalahan yang dilakukan oleh model. Perhitungan evaluasi metrik *accuracy*, *precision*, *recall*, dan *F1-score* dapat dilihat sebagai berikut.

1. **Accuracy**

Accuracy mengukur ketepatan model klasifikasi dengan membandingkan jumlah prediksi benar terhadap total data yang diuji [63].

Rumus 2.7 *Accuracy*.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.7)$$

2. **Precision**

Precision mengukur proporsi prediksi positif yang benar terhadap seluruh prediksi positif yang dihasilkan oleh model [63].

Rumus 2.8 *Precision*.

$$Precision = \frac{TP}{TP + FP} \quad (2.8)$$

3. **Recall**

Recall mengukur proporsi prediksi positif yang benar dibandingkan dengan seluruh kasus positif aktual [63].

Rumus 2.9 *Recall*.

$$Recall = \frac{TP}{TP + FN} \quad (2.9)$$

4. **F1-Score**

F1-Score merupakan rata-rata harmonis yang dibobotkan dari *precision* dan *recall*, sehingga memberikan gambaran seimbang antara keduanya [63].

Rumus 2.10 *F1-Score*.

$$F - 1score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.10)$$

