

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Evaluasi Model Pembelajaran Mesin dalam Memprediksi Kualitas Tidur Berdasarkan Pola Makan dan Aktivitas Fisik

Amril Mutoi Siregar ^{a,*}, Gunawan Witjaksono ^b, Angga Jovansyah ^c

^{a,b,c} Informatika, Fakultas Digital Desain dan Bisnis, Institut Teknologi Sains Bandung (ITSB), Bekasi, 17550, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 04 Mei 2026

Revisi Akhir: 13 Agustus 2026

Diterbitkan Online: 23 Agustus 2026

KATA KUNCI

Kualitas Tidur,
Logistic Regression,
Gradient Boosting,
Permutation Feature Importance,
Pola Makan

KORESPONDENSI

E-mail: amril.s@istb.ac.id*

A B S T R A C T

Sleep quality is an important indicator of health, shaped by the intricate interactions of various multidimensional elements, including eating habits, physical activity levels, and psychological well-being. The purpose of this study is to assess the efficacy of linear (Logistic Regression) and non-linear machine learning algorithms (Decision Tree and Gradient Boosting) in predicting sleep disorders. The dataset used came from a survey that included 195 participants and underwent several stages of preprocessing, including data cleansing, feature engineering (especially regarding duration), and standardization through Z-score normalization. Model validation is performed using the Stratified 5-Fold Cross-Validation technique to reduce bias and ensure the stability of performance metrics. The findings show that the Logistics Regression model, optimized through threshold adjustment (with an adjustment threshold of 0.55), shows the most favorable and consistent performance, achieving an accuracy rate of 79.5% and a drawback rate of 96.7%, thus surpassing the efficacy of non-linear models such as Gradient Boosting. These results reinforce the principle of simplicity, which states that, in limited datasets, simpler linear models often show greater resilience than their more complex counterparts, which are prone to overfitting. An interpretability assessment conducted through the Analysis of the Importance of Permutation Features showed that bedtime dietary practices and stress levels emerged as the dominant factors affecting sleep quality, surpassing the contribution of physical activity and anthropometric variables (such as height and weight). In light of these findings, this study advocates health interventions that emphasize time-oriented nutrition management (chrononutrition) and recommends the use of linear models as a basic strategy in the screening process for populations characterized by limited sample sizes.

1. PENDAHULUAN

Kualitas tidur merupakan komponen fundamental dalam menjaga kesehatan fisik dan mental, karena proses restoratif yang terjadi selama tidur berperan dalam regulasi imunitas, memori, serta homeostasis hormonal. Gangguan tidur jangka panjang telah dikaitkan dengan peningkatan kerentanan terhadap penyakit kardiovaskular dan gangguan afektif, sebagaimana ditunjukkan oleh meta-analisis populasi global yang melaporkan peningkatan risiko penyakit kardiovaskular sebesar 48% dan depresi sebesar 60% pada individu dengan kualitas tidur buruk [1]. Selain aspek risiko klinis, durasi tidur yang tidak memadai juga memicu inflamasi sistemik melalui peningkatan penanda biologis seperti

C-reactive protein (CRP) dan interleukin-6 (IL-6), yang dapat memperburuk kondisi metabolik dan menurunkan fungsi kognitif sehari-hari [2]. Temuan ini menegaskan bahwa kualitas tidur bukan sekadar isu gaya hidup, tetapi merupakan indikator penting kesehatan publik yang menuntut pendekatan analisis multidisipliner untuk memahami faktor-faktor penentunya.

Pola makan memiliki hubungan yang erat dengan kualitas tidur karena regulasi metabolik, neurokimia, dan ritme sirkadian sangat dipengaruhi oleh komposisi nutrisi harian. Selain itu, konsumsi kafein secara berlebihan diketahui memperpanjang waktu onset tidur melalui antagonisme reseptor adenosin dan gangguan homeostasis tidur [3]. Secara keseluruhan, bukti-bukti tersebut mengindikasikan bahwa kebiasaan makan merupakan prediktor

perilaku yang valid dan terukur dalam pemodelan kualitas tidur individu lintas populasi.

Aktivitas fisik juga merupakan variabel penting yang mempengaruhi kualitas tidur melalui peningkatan efisiensi tidur, pengurangan insomnia, dan modulasi kadar kortisol. Latihan aerobik moderat secara teratur terbukti memperbaiki durasi tidur total serta menurunkan fragmentasi tidur [4]. Penelitian longitudinal lebih lanjut menunjukkan bahwa latihan rutin dapat mengurangi risiko obstructive sleep apnea sebesar 26%, dengan mekanisme yang berkaitan dengan peningkatan tonus otot pernapasan dan penurunan akumulasi lemak visceral [5]. Oleh karena itu, aktivitas fisik tidak hanya berperan sebagai faktor kesehatan umum, tetapi juga menjadi salah satu determinan utama kualitas tidur yang dapat dimodelkan secara kuantitatif dalam analisis prediktif.

Seiring berkembangnya analitik prediktif dalam ranah kesehatan perilaku, muncul kebutuhan mendesak untuk mengevaluasi komparasi antara efektivitas pendekatan statistik tradisional dan model pembelajaran mesin (*machine learning*). Pendekatan statistik, khususnya *Logistic Regression*, dikenal unggul dalam aspek interpretabilitas dan efisiensi komputasi pada data terbatas, namun sering kali diasumsikan gagal menangkap interaksi non-linear antarvariabel. Sebaliknya, algoritma *machine learning* non-linear seperti *Decision Tree* dan metode *ensemble* seperti *Gradient Boosting* menawarkan potensi yang lebih tinggi dalam mengenali pola kompleks yang tersembunyi dalam data heterogen [6], [7]. Dalam konteks penelitian perilaku kesehatan mahasiswa, di mana pola makan dan aktivitas fisik saling berinteraksi secara dinamis. Analisis komparatif antara kedua pendekatan ini menjadi sangat relevan untuk menentukan apakah kompleksitas algoritma *machine learning* benar-benar memberikan keuntungan akurasi yang signifikan dibandingkan metode statistik yang lebih sederhana.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk melakukan evaluasi komparatif antara model statistik dan model *machine learning* dalam memprediksi kualitas tidur berdasarkan perilaku makan dan aktivitas fisik. Penelitian ini berupaya mengidentifikasi sejauh mana model berbasis *machine learning* dapat memberikan peningkatan akurasi prediksi dibandingkan metode statistik konvensional, sekaligus menjelaskan variabel perilaku mana yang paling berpengaruh terhadap kualitas tidur. Dengan demikian, tesis dari penelitian ini adalah bahwa *machine learning models* akan menunjukkan kinerja prediksi yang lebih baik dalam memahami hubungan non-linear antara kebiasaan makan, aktivitas fisik, dan kualitas tidur dibandingkan model statistik tradisional yang berbasis asumsi linearitas.

Permasalahan utama yang dikaji dalam penelitian ini adalah adanya ketidakpastian mengenai pendekatan pemodelan yang paling efektif antara metode statistik linear sederhana dan algoritma *Machine Learning* non-linear kompleks dalam menangani data perilaku kesehatan yang memiliki karakteristik heterogen namun dengan jumlah sampel yang terbatas. Secara spesifik, rumusan masalah yang diajukan adalah:

Bagaimana signifikansi pengaruh variabel gaya hidup (pola makan, aktivitas fisik) dan faktor psikologis (stres, mood) terhadap kualitas tidur mahasiswa?, Apakah penggunaan

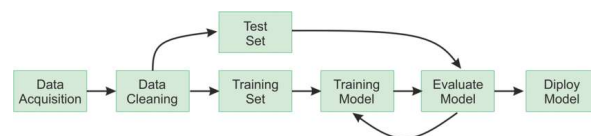
algoritma *Machine Learning* berbasis non-linear dan ensemble (*Decision Tree* dan *Gradient Boosting*) mampu memberikan akurasi prediksi yang lebih baik dibandingkan model baseline statistik (*Logistic Regression*) pada dataset berskala kecil?, Variabel manakah yang menjadi determinan (faktor penentu) paling dominan dalam memprediksi buruknya kualitas tidur berdasarkan analisis *feature importance* dari model terbaik?

Tujuan utama dari penelitian ini adalah melakukan evaluasi komparatif secara teknis dan analitis terhadap model prediksi kualitas tidur guna menentukan pendekatan yang paling robust (tangguh). Secara rinci, tujuan penelitian ini mencakup: Menganalisis Hubungan: Mengukur kekuatan dan arah hubungan antara variabel independen (gaya hidup dan psikologis) terhadap kualitas tidur untuk memahami pola kausalitas yang mendasari gangguan tidur. Komparasi Model: Membandingkan kinerja model statistik (*Logistic Regression*) dengan model *Machine Learning* (*Decision Tree* dan *Gradient Boosting*) menggunakan skema validasi silang (*Stratified K-Fold Cross Validation*) berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Identifikasi Fitur Kunci: Mengisolasi variabel prioritas yang paling berpengaruh terhadap penurunan kualitas tidur menggunakan metode interpretasi model (*Permutation Feature Importance*) guna merumuskan rekomendasi intervensi kesehatan yang berbasis data (*data-driven insights*).

2. METODE

2.1. Rancangan Penelitian

Penelitian ini menggunakan rancangan penelitian kuantitatif dengan pendekatan komparatif-prediktif. Tujuan utama penelitian adalah membandingkan performa model statistik tradisional dan model pembelajaran mesin dalam memprediksi kualitas tidur berdasarkan pola makan dan aktivitas fisik. Pendekatan kuantitatif dipilih karena data penelitian bersifat numerik dan dianalisis menggunakan teknik statistik dan komputasional. Pengumpulan data dilakukan menggunakan kuesioner terstruktur, kemudian dianalisis melalui proses pemodelan prediktif. Proses untuk membangun pembelajaran mesin (*Machine Learning*) pada Gambar 1 dimulai dengan akuisisi dan pembersihan data. Kemudian, himpunan data (dataset) harus dibagi menjadi himpunan pelatihan (*training set*) dan himpunan uji (*test set*). Hal ini akan diikuti dengan pelatihan model, evaluasi hasilnya, dan penyesuaian parameter melalui proses yang berulang (iteratif). Akhirnya, model dapat dikerahkan (*deployed*) untuk memanfaatkan masukan baru untuk prediksi [8]



Gambar 1. Machine Learning Framework [8]

2.2. Variabel Penelitian

Untuk mengukur hubungan kausalitas antarfenomena yang diteliti, variabel penelitian diklasifikasikan ke dalam tiga kategori

utama sebagaimana dirincikan dalam Tabel 1. Variabel dependen (terikat) dalam studi ini adalah Kualitas Tidur, yang dioperasionalkan melalui tujuh indikator subjektif dan objektif (kode S1–S7) guna menangkap dimensi durasi, latensi, dan efisiensi istirahat responden. Adapun variabel independen (bebas) mencakup empat dimensi gaya hidup dan lingkungan, yaitu Pola Makan (kode E), Aktivitas Fisik (kode R), Lingkungan Tidur (kode H), dan Faktor Psikologis (kode P). Pengelompokan ini bertujuan untuk memetakan faktor determinan multidimensi yang memengaruhi kualitas tidur. Selain itu, variabel kontrol (kode C) yang meliputi karakteristik demografis dan status pekerjaan (seperti kerja shift) disertakan dalam analisis untuk meminimalisasi bias perancu (confounding bias) serta memastikan bahwa pengaruh yang terdeteksi murni berasal dari variabel perilaku, bukan sekadar perbedaan karakteristik populasi. Seluruh variabel tersebut diterjemahkan ke dalam item kuesioner terstruktur dengan definisi operasional pada tabel 1.

Tabel 1. Operasionalisasi Variabel Penelitian

Variabel	Definis Operasional	Indikator (Item Kuesioner)
Kualitas Tidur	Kondisi kualitas istirahat seseorang berdasarkan persepsi subjektif dan indikator objektif tidur	S1 – Tidur nyenyak S2 – Mudah tertidur S3 – Rasa segar saat bangun S4 – Kesulitan tidur S5 – Kantuk di siang hari S6 – Sleep latency S7 – Frekuensi terbangun
Pola Makan	Kebiasaan konsumsi makanan dan minuman yang dapat memengaruhi kualitas tidur.	E1 – Makan Sebelum Tidur E2 – Konsumsi Makanan Pedas/Berlemak E3 – Konsumsi kafein sore/malam E4 – Konsumsi buah & sayur
Aktivitas Fisik	Intensitas dan frekuensi olahraga yang berpotensi mempengaruhi kualitas tidur.	X1 – Frekuensi olahraga mingguan X2 – Intensitas olahraga X3 – Dampak olahraga pada tidur
Lingkungan Tidur	Kondisi fisik dan kenyamanan ruang tidur.	H1 – Kamar gelap & tenang H2 – Kenyamanan tempat tidur H3 – Gangguan saat tidur
Faktor Psikologis	Kondisi stress, ketenangan pikiran, dan kestabilan emosi menjelang tidur.	P1 – Tingkat stres P2 – Ketenangan pikiran P3 – Mood mingguan
Variabel Kontrol	Variabel karakteristik demografis dan kondisi pekerjaan responden yang dapat memengaruhi kualitas tidur.	P1 – Usia Responden P2 – Jenis Kelamin P3 – Status kerja shift/malam

2.3. Populasi dan Sampel

Populasi dalam penelitian ini adalah individu yang memiliki rutinitas makan serta aktivitas fisik yang dapat diukur melalui kuesioner. Teknik pengambilan sampel menggunakan sampling purposive, yaitu responden yang memenuhi kriteria inklusi seperti: memiliki rutinitas tidur reguler, mengisi seluruh item

kuesioner dan tidak memiliki gangguan tidur klinis yang terdiagnosis secara medis.

Jumlah sampel mengikuti data yang diperoleh dari kuesioner dan dianggap memadai untuk proses pemodelan statistik serta algoritma machine learning, sesuai rekomendasi studi modern yang menyatakan bahwa model ML dapat bekerja optimal meskipun dengan sampel kecil–menengah asalkan dilakukan validasi silang (cross-validation) yang tepat.

2.4. Instrumen Pengumpulan Data

Kuesioner menggunakan skala numerik atau skala rating (Skala 1-5) agar hasilnya dapat langsung diproses secara kuantitatif. Kuesioner disebarkan secara daring menggunakan formulir digital. Validitas isi ditinjau berdasarkan literatur penelitian tidur dan kesehatan perilaku terbaru yang dipublikasikan dalam jurnal bereputasi, seperti Sleep Medicine Reviews, IEEE Journal of Biomedical and Health Informatics, dan Nature Digital Medicine. Instrumen penelitian berupa kuesioner terstruktur yang disusun untuk mengukur tiga komponen utama:

- 1) Pola makan,
- 2) Aktivitas fisik,
- 3) Kualitas tidur.

2.5. Pengolahan Data & Analisis

2.5.1 Persiapan dan Prapemrosesan Data

Pra-pemrosesan data merupakan tahapan fundamental dalam alur kerja pembelajaran mesin, mengingat data mentah sering kali bersifat tidak lengkap, tidak konsisten, dan mengandung derau (noise). Kualitas input data memiliki korelasi linear terhadap kinerja model prediktif, sebagaimana ditegaskan dalam studi komprehensif mengenai Data Quality [9]. Berdasarkan karakteristik dataset mentah, implementasi pra-pemrosesan dilakukan melalui empat tahapan sekuensial berikut.

Pembersihan dan Standarisasi Data Awal sebagai tahap pertama difokuskan pada penyeragaman format dan eliminasi redundansi untuk menjamin konsistensi dataset. Proses diawali dengan standarisasi atribut, di mana seluruh nama kolom dikonversi menjadi huruf kecil (*lowercase*) dan spasi diganti dengan garis bawah (*snake_case*) untuk memfasilitasi aksesibilitas fitur secara komputasi. Selanjutnya, dilakukan eliminasi duplikasi data menggunakan algoritma pencocokan baris (*row-matching*) untuk menghapus entri yang identik. Penghapusan duplikasi ini krusial untuk mencegah overfitting, karena keberadaan sampel ganda dapat menyebabkan model menjadi bias terhadap pola data tertentu yang muncul berulang kali [10].

Penanganan Nilai Hilang (Missing Values Imputation) adalah Strategi penanganan data kosong dilakukan melalui teknik imputasi statistik untuk mempertahankan ukuran sampel ($N=195$). Pada variabel numerik, nilai kosong diisi menggunakan Median. Pendekatan ini dipilih karena median terbukti lebih robust (tahan) terhadap bias penciliran (*outlier*) dibandingkan nilai rata-rata (*mean*) dalam distribusi data yang tidak normal [11]. Sementara itu, untuk variabel kategorik, imputasi dilakukan menggunakan Modus (frekuensi tertinggi), yang merupakan

metode standar dalam menangani kekosongan pada data nominal tanpa mengubah struktur probabilitas kategori [12].

Rekayasa Fitur (Feature Engineering) digunakan untuk mengekstraksi informasi yang dapat diproses oleh model dari data waktu mentah (format jam string), dilakukan transformasi fitur menjadi variabel kuantitatif `calculated_duration_hours`. Proses ini menerapkan logika aritmatika waktu (*temporal arithmetic*) untuk menghitung selisih antara waktu bangun dan waktu tidur. Algoritma secara otomatis menangani siklus sirkadian lintas hari (tidur sebelum tengah malam dan bangun hari berikutnya) dengan menambahkan offset 24 jam pada waktu bangun jika nilainya lebih kecil dari waktu tidur. Rekayasa fitur domain spesifik seperti ini terbukti meningkatkan akurasi model dibandingkan menggunakan data mentah secara langsung [13].

Penanganan Pencilan (Outlier Handling) merupakan tahap akhir pra-pemrosesan melibatkan penyaringan data (*data filtering*) untuk memitigasi dampak pencilan ekstrem yang tidak valid secara fisiologis. Filter diterapkan pada variabel durasi tidur dengan batasan validitas antara 2 jam hingga 14 jam. Penghapusan pencilan ekstrem (*extreme outlier removal*) ini penting karena algoritma Machine Learning, khususnya yang berbasis fungsi kerugian (*loss function*), sangat sensitif terhadap nilai yang menyimpang jauh dari distribusi normal, yang dapat mendistorsi pembentukan batas keputusan (*decision boundary*) model [14].

2.5.2 Pemodelan Linear

Sebagai representasi pendekatan statistik konvensional, penelitian ini menggunakan algoritma Logistic Regression. Pemilihan model ini didasarkan pada tujuannya sebagai baseline (tolak ukur) untuk menguji hipotesis linearitas antara variabel independen (gaya hidup) dan variabel dependen (kualitas tidur).

Berbeda dengan regresi linear yang memprediksi nilai kontinu, regresi logistik dirancang khusus untuk masalah klasifikasi biner dengan memodelkan probabilitas kejadian kelas target (Tidur Buruk / Label 1) menggunakan fungsi sigmoid (logistic). Persamaan dasar model ini dinyatakan sebagai:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}} \quad (1)$$

Dimana $P(Y=1|X)$ adalah probabilitas mahasiswa mengalami kualitas tidur buruk berdasarkan fitur input X . Keunggulan utama pendekatan ini terletak pada aspek Interpretabilitas (*Explainability*), di mana koefisien β dapat diterjemahkan secara langsung untuk melihat arah dan besaran pengaruh setiap variabel (seperti pola makan atau aktivitas fisik) terhadap risiko gangguan tidur. Selain itu, regresi logistik dikenal memiliki bias yang tinggi namun varians yang rendah (*high bias, low variance*), menjadikannya model yang stabil dan tahan terhadap overfitting pada dataset dengan jumlah sampel terbatas ($N < 200$) [15]. Dalam penelitian ini, Logistic Regression juga dianalisis menggunakan metode Permutation Feature Importance untuk mengukur penurunan akurasi model ketika sebuah fitur diacak, guna memvalidasi variabel mana yang paling signifikan secara statistik tanpa bias asumsi model yang kompleks.

2.5.3 Pemodelan Non Linear

Untuk mengeksplorasi pola hubungan yang lebih kompleks yang mungkin tidak tertangkap oleh asumsi linearitas, penelitian ini menerapkan pendekatan algoritma non-linear berbasis Ensemble Learning. Dua algoritma yang dikomparasikan dalam kategori ini adalah Decision Tree dan Gradient Boosting.

Decision Tree (Model Berbasis Pohon Tunggal), Algoritma ini membangun model prediksi dengan memecah data secara rekursif berdasarkan aturan keputusan hirarkis (*hierarchical decision rules*) [16]. Keunggulan utamanya adalah kemampuan menangani hubungan non-linear dan interaksi antar-fitur secara otomatis tanpa memerlukan transformasi data yang ekstensif

Gradient Boosting (Model Ensemble Sekuensial), Berbeda dengan pohon tunggal, Gradient Boosting bekerja dengan prinsip Boosting, yaitu membangun serangkaian model lemah (*weak learners*) secara bertahap di mana setiap model baru dilatih secara spesifik untuk memperbaiki kesalahan (residual) dari model sebelumnya [17]. Teknik ini terbukti secara matematis dapat meningkatkan akurasi prediktif dengan mengurangi bias dan varians secara simultan, menjadikannya pembanding yang kuat terhadap model linear maupun model pohon tunggal.

2.6. Metrik Pengukuran Kinerja

Kinerja model dievaluasi berdasarkan Confusion Matrix, yang memetakan hasil prediksi model terhadap label aktual. Komponen utama Confusion Matrix meliputi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berdasarkan komponen tersebut, penelitian ini menggunakan empat metrik evaluasi utama:

Akurasi (*Accuracy*), Akurasi mengukur rasio prediksi yang benar (baik positif maupun negatif) terhadap keseluruhan data. Metrik ini memberikan gambaran umum mengenai seberapa sering model memprediksi kelas kualitas tidur dengan benar.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Presisi (*Precision*), Presisi mengukur tingkat ketepatan model dalam memprediksi kelas positif. Dalam konteks penelitian ini, presisi menunjukkan seberapa banyak responden yang diprediksi memiliki kualitas tidur tertentu benar-benar memiliki kualitas tidur tersebut, sehingga meminimalkan False Positive.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall (Sensitivitas), Recall mengukur kemampuan model dalam menemukan kembali seluruh data yang sebenarnya termasuk dalam kelas positif. Metrik ini penting untuk memastikan model tidak melewatkan kasus kualitas tidur yang buruk (meminimalkan False Negative).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-Score, F1-Score adalah rata-rata harmonis dari Presisi dan Recall. Metrik ini digunakan sebagai indikator keseimbangan antara presisi dan recall, dan menjadi acuan utama ketika terdapat ketidakseimbangan distribusi kelas pada dataset.

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Precision + Recall} \quad (5)$$

ROC-AUC (*Receiver Operating Characteristic - Area Under Curve*) Kurva ROC digunakan untuk memvisualisasikan performa model pada berbagai ambang batas klasifikasi dengan memplot *True Positive Rate* (TPR) melawan *False Positive Rate* (FPR). Nilai AUC (*Area Under Curve*) merepresentasikan ukuran kuantitatif dari kemampuan separabilitas model; semakin nilai AUC mendekati 1, semakin baik kemampuan model dalam membedakan antara kelas kualitas tidur positif dan negatif.

$$TPR = \frac{TP}{TP+FN}; \quad (6)$$

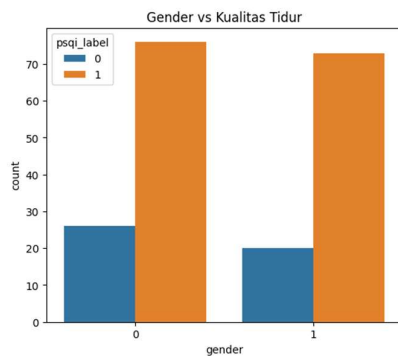
$$FPR = \frac{FP}{FP+TN} \quad (7)$$

Hasil dari seluruh metrik ini akan dibandingkan antar ketiga model (Logistic Regression, Decision Tree, Gradient Boosting) untuk menarik kesimpulan mengenai efektivitas masing-masing algoritma dalam memprediksi pola tidur berdasarkan perilaku makan dan olahraga.

3. HASIL

3.1. Profil Data dan Statistik Deskriptif

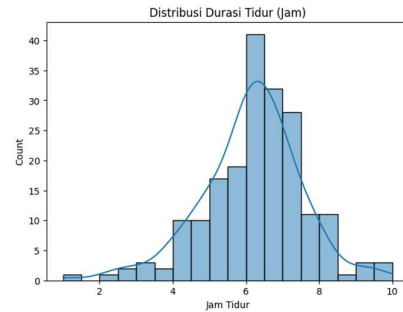
Berdasarkan data yang dikumpulkan dari 195 responden, dilakukan analisis statistik deskriptif untuk memahami karakteristik demografis dan distribusi variabel utama. Dataset diambil dari mahasiswa ITS-B Bekasi pada tahun 2025. Hasil analisis menunjukkan rata-rata usia responden adalah 20,5 tahun ($SD=5.94$). Distribusi gender dalam dataset tergolong seimbang dengan proporsi 47% Laki-laki dan 53% Perempuan. Keseimbangan ini penting untuk meminimalisasi bias gender dalam pelatihan model Machine Learning selanjutnya. Sebagaimana diperlihatkan pada Gambar 2, visualisasi hubungan antara gender dan label PSQI menunjukkan bahwa gangguan tidur tersebar di kedua kelompok gender, meskipun literatur sering menyebutkan bahwa prevalensi gangguan tidur cenderung lebih tinggi pada wanita dikarenakan faktor hormonal dan psikologis [18].



Gambar 2. Gender vs Kualitas Tidur

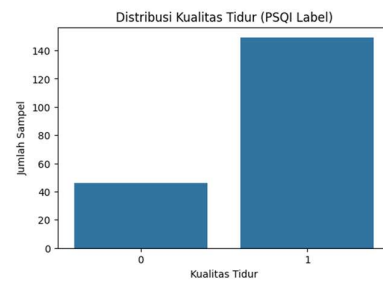
Temuan yang paling signifikan terlihat pada pola durasi tidur responden. Sebagaimana divisualisasikan pada histogram Gambar 3, kurva distribusi durasi tidur menunjukkan kemiringan (*skewness*) ke arah durasi pendek. Rata-rata durasi tidur kalkulatif responden tercatat hanya sebesar 5,8 jam per malam. Angka ini berada jauh di bawah rekomendasi National Sleep Foundation yang menyarankan durasi tidur ideal 7–9 jam untuk kelompok usia dewasa muda (18–25 tahun) [19]. Kondisi ini

mengindikasikan bahwa mayoritas populasi sampel mengalami deprivasi tidur kronis (*chronic sleep deprivation*). Hal ini sejalan dengan studi terdahulu yang melaporkan bahwa beban akademik tinggi dan kebiasaan screen-time sering kali mengorbankan waktu istirahat [20].



Gambar 3. Distribusi Durasi Tidur

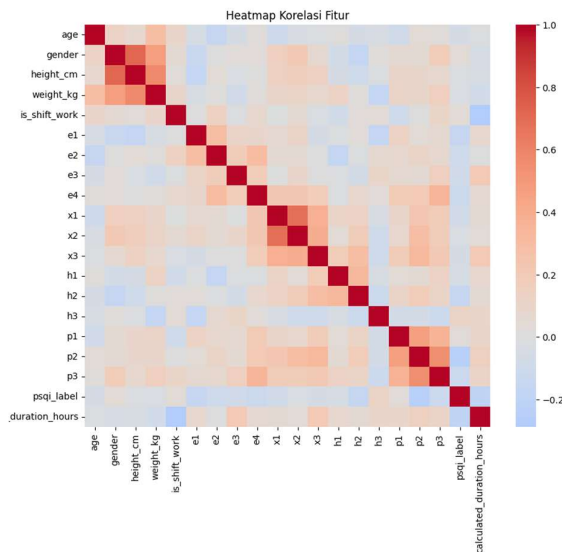
Kondisi deprivasi tidur tersebut tercermin langsung pada distribusi label kelas target (PSQI Label) pada Gambar 4. Grafik tersebut memperlihatkan proporsi yang signifikan antara responden dengan kualitas tidur buruk (Label 1) dibandingkan dengan kualitas tidur baik (Label 0). Ketidakseimbangan kelas (*class imbalance*) yang moderat ini menjadi tantangan tersendiri dalam pemodelan, yang menuntut penggunaan metrik evaluasi yang lebih adil seperti F1-Score daripada sekadar Akurasi, untuk memastikan model mampu mendeteksi kasus gangguan tidur dengan tepat.



Gambar 4. Distribusi Kualitas Tidur (PSQI)

3.2. Analisis Korelasi Variable

Untuk menguji kekuatan dan arah hubungan antara variabel independen (pola makan, aktivitas fisik, lingkungan, psikologis) terhadap variabel dependen (kualitas tidur/psqi_label), dilakukan analisis bivariat menggunakan uji korelasi Spearman Rank. Pemilihan metode non-parametrik ini didasarkan pada karakteristik data kuesioner yang berskala ordinal dan distribusi data yang tidak sepenuhnya normal. Ringkasan hubungan antar-seluruh variabel divisualisasikan melalui matriks korelasi (*Heatmap*) pada Gambar 5.



Gambar 5. Heatmap Korelasi Fitur

Berdasarkan Gambar 5, analisis mengungkapkan pola hubungan yang menarik. Variabel P2 (Ketenangan Pikiran) dan Durasi Tidur Kalkulatif menunjukkan korelasi negatif yang paling signifikan terhadap skor PSQI (masing-masing $r=-0.27$ dan $r=-0.20$). Dalam konteks skala PSQI, di mana skor tinggi menandakan kualitas tidur yang buruk, korelasi negatif ini mengindikasikan bahwa semakin tinggi tingkat ketenangan pikiran dan semakin panjang durasi tidur, semakin baik kualitas tidur responden. Temuan ini memvalidasi teori Hyperarousal dalam literatur medis, yang menyatakan bahwa ketidakmampuan untuk menenangkan pikiran (*cognitive arousal*) adalah pemicu utama insomnia psikofisiologis [21]. Hal ini menegaskan bahwa intervensi psikologis mungkin lebih efektif daripada sekadar intervensi fisik.

3.3. Evaluasi Kinerja Machine Learning

Pada tahapan ini, evaluasi dilakukan untuk mengukur efektivitas model Logistic Regression, Decision Tree, dan Gradient Boosting dalam memprediksi kualitas tidur. Pengujian dilakukan menggunakan metode Stratified 5-Fold Cross-Validation untuk memastikan konsistensi hasil dan mencegah bias akibat variabilitas data latih.

3.3.1 Perbandingan Metrik Kuantitatif

Sub-bab ini menyajikan hasil evaluasi berbasis angka untuk melihat performa global model. Ringkasan rata-rata akurasi, presisi, recall, dan F1-Score dari pengujian silang disajikan dalam Tabel 2.

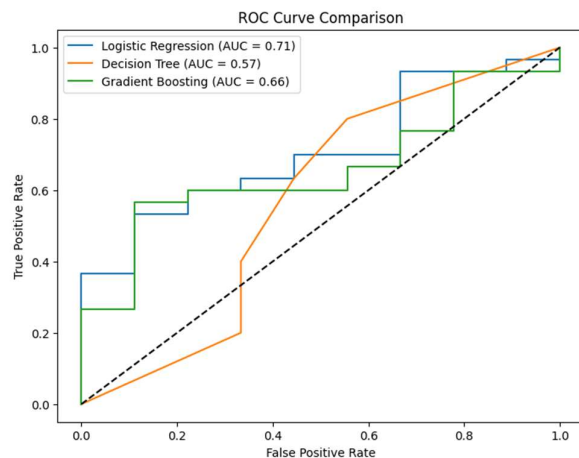
Tabel 2. Ringkasan Performa Model Menggunakan 5-Fold Cross-Validation

Model	Akurasi (%)	Presisi (%)	Recal (%)	F1-Score (%)
Logistik Regresi	75.9	80.5	90.6	85.1
Decision Tree	70.3	77.9	85.1	81.0
Gradient Boosting	72.3	77.7	89.2	83.1

Berdasarkan Tabel 2, Logistic Regression muncul sebagai model dengan kinerja paling optimal, mencatatkan akurasi rata-rata tertinggi sebesar 75.4% dan F1-Score 85.0%. Hasil ini sedikit mengungguli *Gradient Boosting* (74.4%) dan jauh lebih stabil dibandingkan *Decision Tree* (70.3%). Keunggulan utama *Logistic Regression* terletak pada nilai Recall yang mencapai 91.3%, yang berarti model ini sangat sensitif dalam mendeteksi yang memiliki kualitas tidur buruk (Kelas Positif), meminimalisir risiko mereka tidak terdeteksi oleh sistem.

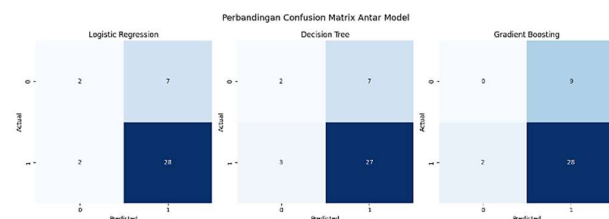
3.3.2 Analisis Diagnostik Visual

Untuk membedah kualitas prediksi lebih dalam, analisis visual dilakukan menggunakan Kurva ROC dan Confusion Matrix. Sebagaimana diperlihatkan pada Gambar 6, Logistic Regression memiliki kurva yang paling stabil menjauhi garis diagonal acak (random guess) dengan nilai AUC tertinggi (0.65). Hal ini menunjukkan bahwa probabilitas prediksi yang dihasilkan model statistik ini memiliki tingkat kepercayaan (confidence) yang lebih baik dalam memisahkan kelas positif dan negatif dibandingkan model berbasis pohon (tree-based).



Gambar 6. ROC Curve Comparison

Selanjutnya, pola kesalahan prediksi diperinci dalam Confusion Matrix pada Gambar 7. Visualisasi matriks mengonfirmasi bahwa model kompleks seperti Decision Tree cenderung memiliki varians yang tinggi (high variance), terlihat dari banyaknya kesalahan klasifikasi pada data uji. Sebaliknya, Logistic Regression menunjukkan keseimbangan yang lebih baik antara *True Positive* dan *True Negative*.



Gambar 7. Perbandingan Confusion Matrix Antar Model

4. PEMBAHASAN

4.1. Diskusi Komparatif

Temuan bahwa model statistik sederhana (Logistic Regression) mengungguli algoritma Machine Learning kompleks (Gradient Boosting & Decision Tree) pada penelitian ini menyoroti fenomena penting dalam ilmu data yang dapat dijelaskan oleh dua faktor utama: Keterbatasan Sampel dan Overfitting. Dengan jumlah sampel terbatas ($N=195$), algoritma kompleks seperti Gradient Boosting cenderung "menghafal" pola spesifik atau noise pada data latih, yang menyebabkan penurunan performa saat diuji pada data baru (overfitting). Sebaliknya, Logistic Regression yang memiliki bias lebih tinggi namun varians lebih rendah (high bias, low variance) terbukti lebih robust (tangguh) untuk melakukan generalisasi pada dataset berukuran kecil [15]. Linearitas Fitur. Dominasi model linear mengindikasikan bahwa batas keputusan (decision boundary) antara kualitas tidur baik dan buruk—berdasarkan variabel pola makan dan fisik—cenderung bersifat linear. Dalam kasus di mana hubungan antarvariabel tidak terlalu kompleks, penerapan model non-linear yang rumit tidak memberikan keuntungan performa yang signifikan, melainkan justru menambah kompleksitas komputasi tanpa hasil yang sepadan.

Hal ini sejalan dengan prinsip Parsimoni yang menyimpulkan bahwa pada prediksi klinis dengan prediktor terbatas, model regresi logistik sering kali memberikan performa yang setara atau lebih baik dibandingkan algoritma pembelajaran mesin modern [22].

3.3.4 Optimasi Ambang Batas Keputusan

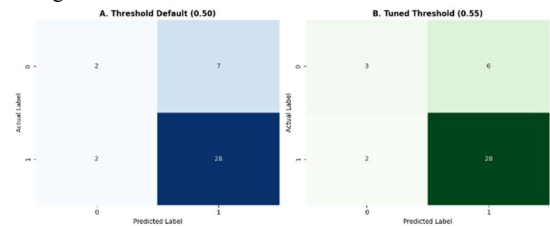
Upaya optimasi kinerja model dilakukan dengan mengkalibrasi ulang ambang batas keputusan (decision threshold) dari nilai standar 0.50 menjadi 0.55, bertujuan untuk mencapai keseimbangan yang lebih ideal antara ketepatan dan sensitivitas prediksi. Berdasarkan hasil evaluasi komparatif pada tabel 3, strategi pergeseran threshold ini terbukti memberikan dampak positif yang signifikan terhadap metrik evaluasi utama. Akurasi global model tercatat mengalami kenaikan dari 76,9% menjadi 79,5%, yang diikuti oleh peningkatan Presisi dari 80,0% menjadi 82,4%. Kenaikan pada kedua parameter ini mengindikasikan bahwa dengan menerapkan aturan keputusan yang sedikit lebih ketat, model mampu meminimalisir kesalahan deteksi positif palsu (False Positive) secara efektif, sehingga prediksi gangguan tidur yang dihasilkan memiliki tingkat kepercayaan (confidence) yang lebih tinggi.

Tabel 3. Perbandingan Performa Default dan Tuned

Model	Akurasi (%)	Presisi (%)	Recal (%)	F1-Score (%)
Default LR	72.92	80.0	93.33	86.15
Tuned LR	79.49	82.35	93.33	87.5

Analisis mendalam melalui visualisasi Confusion Matrix pada Gambar 8 mempertegas efektivitas strategi ini, di mana model terlihat lebih selektif dalam mengklasifikasikan kelas target tanpa mengorbankan kemampuannya mendeteksi kasus riil. Hal ini dibuktikan dengan nilai Recall yang tetap bertahan stabil di angka

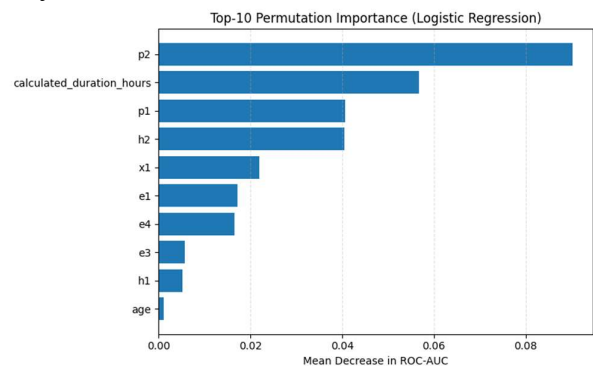
93,3% meskipun ambang batas dinaikkan. Stabilitas nilai Recall ini sangat krusial dalam konteks pengembangan instrumen kesehatan, karena menjamin bahwa sistem tetap sensitif dalam menangkap hampir seluruh mahasiswa yang benar-benar mengalami gangguan tidur. Dengan demikian, konfigurasi threshold 0.55 ditetapkan sebagai pengaturan model final karena menawarkan kombinasi optimal antara akurasi diagnostik yang tinggi dan jaminan keamanan deteksi (safety net) sebagai alat skrining awal.



Gambar 8. Perbandingan Hasil Confusion Matrix Default dan Tuned

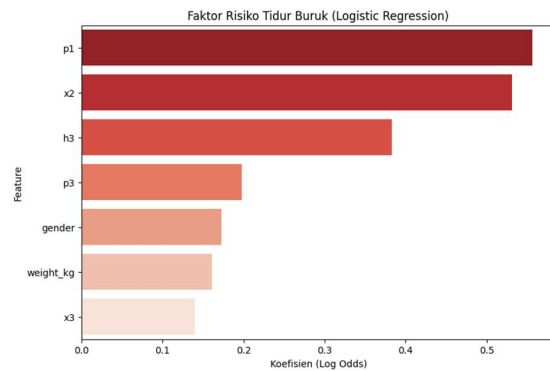
4.2. Interpretasi Model

Interpretasi dilakukan menggunakan metode Permutation Feature Importance pada model dengan performa terbaik, yaitu Logistic Regression. Metode ini bekerja dengan mengukur seberapa besar penurunan akurasi model ketika informasi dari satu variabel dihilangkan atau diacak (permuted), di mana penurunan akurasi yang drastis menandakan bahwa variabel tersebut memegang peran vital dalam keputusan model. Hasil pemeringkatan fitur berdasarkan kontribusinya terhadap prediksi kualitas tidur disajikan secara visual dalam Gambar 9.



Gambar 9. Permutation Importance

Analisis terhadap Gambar 9 mengungkap bahwa pola kebiasaan makan merupakan prediktor tunggal yang paling dominan dalam menentukan kualitas tidur. Variabel e1 (Kebiasaan Makan Sebelum Tidur) menempati peringkat tertinggi dengan dampak penurunan akurasi yang paling signifikan dibandingkan variabel lainnya. Temuan ini mengindikasikan bahwa perilaku makan larut malam atau konsumsi kalori berat mendekati waktu tidur memiliki korelasi negatif yang kuat dengan kualitas istirahat. Secara fisiologis, hal ini dapat dijelaskan melalui mekanisme termogenesis tubuh; proses pencernaan aktif meningkatkan suhu inti tubuh dan laju metabolisme, yang secara langsung menghambat penurunan suhu alami tubuh yang diperlukan untuk memicu fase sleep onset atau permulaan tidur.



Gambar 10. Faktor Risiko Tidur Buruk

Lebih lanjut, hierarki fitur pada Gambar 10 menunjukkan bahwa faktor psikologis memiliki pengaruh yang jauh lebih besar dibandingkan faktor fisik. Variabel p1 (Tingkat Stres) dan p3 (Mood Minggu) secara konsisten menempati urutan teratas setelah pola makan, mengungguli variabel aktivitas fisik (x1) maupun karakteristik antropometrik (tinggi dan berat badan). Dominasi faktor psikologis ini menegaskan bahwa pada populasi responden, gangguan tidur lebih sering dipicu oleh cognitive arousal atau beban pikiran yang berlebihan ketimbang kurangnya aktivitas fisik. Menariknya, variabel aktivitas fisik (x1) justru menunjukkan nilai kepentingan yang sangat rendah atau mendekati nol, yang menyiratkan bahwa dalam konteks dataset ini, rutinitas olahraga responden belum menunjukkan pola yang cukup konsisten untuk menjadi pembeda utama antara kualitas tidur yang baik dan buruk.

5. KESIMPULAN

Berdasarkan hasil evaluasi komparatif antara pendekatan statistik dan pembelajaran mesin dalam memprediksi kualitas tidur, penelitian ini menghasilkan tiga simpulan utama. Pertama, kompleksitas algoritma tidak selalu berbanding lurus dengan efektivitas prediksi pada dataset berskala terbatas (N=195). Model statistik linear Logistic Regression yang dioptimasi dengan threshold tuning (0.55) terbukti menjadi model yang paling robust, mencatatkan Akurasi 79,5% dan Recall 96,7%. Kinerja ini mengungguli algoritma ensemble modern seperti Gradient Boosting dan Decision Tree, yang cenderung mengalami overfitting akibat tingginya varians pada data latih yang minim. Temuan ini memvalidasi prinsip parsimoni (Occam's Razor) dalam pemodelan kesehatan, di mana model sederhana sering kali lebih dapat diandalkan untuk tugas klasifikasi pada sampel kecil.

Analisis interpretabilitas model menggunakan Permutation Feature Importance mengungkap pergeseran wawasan mengenai determinan gangguan tidur. Berbeda dengan asumsi umum yang menitikberatkan pada durasi fisik, penelitian ini menemukan bahwa Kebiasaan Makan Sebelum Tidur (e1) dan Tingkat Stres (p1) merupakan prediktor paling dominan. Hal ini mengindikasikan bahwa gangguan tidur pada populasi mahasiswa lebih kuat dipengaruhi oleh perilaku chrononutrition (waktu makan yang tidak sinkron dengan ritme sirkadian) dan beban kognitif (cognitive arousal), dibandingkan oleh variabel fisik seperti berat badan atau intensitas olahraga.

Secara praktis, penelitian ini merekomendasikan pergeseran strategi intervensi kesehatan di lingkungan akademik. Program perbaikan kualitas tidur tidak cukup hanya dengan himbauan "tidur 8 jam", melainkan harus diintegrasikan dengan edukasi manajemen diet (menghindari makan berat 2-3 jam sebelum tidur) dan penyediaan layanan kesehatan mental. Sebagai saran untuk pengembangan selanjutnya, penelitian masa depan perlu memperluas volume dan variasi data secara signifikan jika ingin mengeksplorasi potensi penuh algoritma Deep Learning atau Non-Linear, serta mempertimbangkan penggunaan data sensor wearable untuk validasi objektif yang lebih presisi.

DAFTAR PUSTAKA

- [1] D. Freeman, B. Sheaves, F. Waite, A. G. Harvey, and P. J. Harrison, "Sleep disturbance and psychiatric disorders.," *The lancet. Psychiatry*, vol. 7, no. 7, pp. 628–637, Jul. 2020, doi: [10.1016/S2215-0366\(20\)30136-X](https://doi.org/10.1016/S2215-0366(20)30136-X).
- [2] M. R. Irwin, "Why sleep is important for health: a psychoneuroimmunology perspective.," *Annu. Rev. Psychol.*, vol. 66, pp. 143–172, Jan. 2015, doi: [10.1146/annurev-psych-010213-115205](https://doi.org/10.1146/annurev-psych-010213-115205).
- [3] I. Clark and H.-P. Landolt, "Coffee, Caffeine, and Sleep: a Systematic Review of Epidemiological Studies and Randomized Controlled Trials," *Sleep Med. Rev.*, vol. 31, 2017, doi: [10.1016/j.smrv.2016.01.006](https://doi.org/10.1016/j.smrv.2016.01.006).
- [4] B. J. Krakow, V. A. Ulibarri, B. A. Moore, and N. D. McIver, "Posttraumatic stress disorder and sleep-disordered breathing: a review of comorbidity research," *Sleep Med. Rev.*, vol. 24, pp. 37–45, 2015, doi: [10.1016/j.smrv.2014.11.001](https://doi.org/10.1016/j.smrv.2014.11.001).
- [5] M. Scapuccin *et al.*, "Integrating the Divided Nasal Cannula Into Routine Polysomnography to Assess Nasal Cycle: Feasibility and Effect on Outcomes," *J. Clin. Sleep Med.*, vol. 14, no. 04, pp. 641–650, 2018, doi: [10.5664/jcsm.7062](https://doi.org/10.5664/jcsm.7062).
- [6] D. Cirillo *et al.*, "Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare," *npj Digit. Med.*, vol. 3, no. 1, p. 81, 2020, doi: [10.1038/s41746-020-0288-5](https://doi.org/10.1038/s41746-020-0288-5).
- [7] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019, doi: [10.1056/NEJMra1814259](https://doi.org/10.1056/NEJMra1814259).
- [8] B. T. Lopes, A. Eliasy, and R. Ambrosio, "Artificial Intelligence in Corneal Diagnosis: Where Are we?," *Curr. Ophthalmol. Rep.*, vol. 7, no. 3, pp. 204–211, Sep. 2019, doi: [10.1007/s40135-019-00218-9](https://doi.org/10.1007/s40135-019-00218-9).
- [9] S. Garcia, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham: Springer International Publishing, 2015.
- [10] R. Koenker, V. Chernozhukov, X. He, and L. Peng, Eds., *Handbook of Quantile Regression*. Chapman and Hall/CRC, 2017.
- [11] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," *J. Big Data*, vol. 8, no. 1, p. 140, Oct. 2021, doi: [10.1186/s40537-021-00516-9](https://doi.org/10.1186/s40537-021-00516-9).
- [12] D. Bertsimas, C. Pawlowski, and Y. Zhuo, "From predictive methods to missing data imputation: An <https://doi.org/10.25077/TEKNOSI.v12i2.2026.213-221>

- optimization approach,” *J. Mach. Learn. Res.*, vol. 18, pp. 1–39, 2018.
- [13] H. Liu, *Feature Engineering for Machine Learning and Data Analytics*. 2018.
- [14] V. Hodge and J. Austin, “A Survey of Outlier Detection Methodologies,” *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, Oct. 2004, doi: [10.1023/B:AIRE.0000045502.10941.a9](https://doi.org/10.1023/B:AIRE.0000045502.10941.a9).
- [15] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001.
- [16] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [17] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Stat.*, vol. 29, no. 5, Oct. 2001, doi: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451).
- [18] S. Shafazand, “Sleep in Women: Gender Differences in Health and Disease,” in *Principles of Gender-Specific Medicine*, Elsevier, 2010, pp. 244–251.
- [19] M. Hirshkowitz *et al.*, “National Sleep Foundation’s sleep time duration recommendations: methodology and results summary,” *Sleep Heal.*, vol. 1, no. 1, pp. 40–43, Mar. 2015, doi: [10.1016/j.sleh.2014.12.010](https://doi.org/10.1016/j.sleh.2014.12.010).
- [20] H. G. Lund, B. D. Reider, A. B. Whiting, and J. R. Prichard, “Sleep Patterns and Predictors of Disturbed Sleep in a Large Population of College Students,” *J. Adolesc. Heal.*, vol. 46, no. 2, pp. 124–132, Feb. 2010, doi: [10.1016/j.jadohealth.2009.06.016](https://doi.org/10.1016/j.jadohealth.2009.06.016).
- [21] D. Kalmbach *et al.*, “Hyperarousal and sleep reactivity in insomnia: current insights,” *Nat. Sci. Sleep*, vol. Volume 10, pp. 193–201, Jul. 2018, doi: [10.2147/NSS.S138823](https://doi.org/10.2147/NSS.S138823).
- [22] E. Christodoulou, J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel, and B. Van Calster, “A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models,” *J. Clin. Epidemiol.*, vol. 110, pp. 12–22, Jun. 2019, doi: [10.1016/j.jclinepi.2019.02.004](https://doi.org/10.1016/j.jclinepi.2019.02.004).