Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 1111~1124
e-ISSN: 2548-6861

Interpretable Ensemble Models for Lifestyle-Based Sleep Disorder Prediction

Farhan Rahardian [1]*, Sindhu Rakasiwi [2]*
* Teknik Informatika, Universitas Dian Nuswantoro
111202214096@mhs.dinus.ac.id [1], sindhu.rakasiwi@dsn.dinus.ac.id [2]

Article Info

Article history:

Received 2025-12-29
Revised 2026-01-26
Accepted 2026-02-11

Keyword:

Ensemble Learning,
Hyperparameter,
Machine Learning,
Sleep Disorder.

ABSTRACT

Sleep disorders are a major global health concern that affect cognitive performance, mental well-being, and long-term physiological health. Conventional diagnostic methods such as polysomnography are time-consuming and resource-intensive, limiting their use for large-scale early detection. Therefore, machine learning offers a practical alternative for predictive and data-driven sleep disorder analysis. This study compares the performance of four ensemble learning algorithms Random Forest, Gradient Boosting, AdaBoost, and XGBoost in predicting sleep disorders based on lifestyle and physiological factors using the Sleep Health and Lifestyle dataset consisting of 374 samples and three classes: insomnia, none, and sleep apnea. The research workflow includes data preprocessing, feature encoding, dataset splitting (70:30), and hyperparameter optimization using Grid Search with 5-fold Cross Validation to improve model stability and generalization given the limited dataset size. Model evaluation is conducted using accuracy, precision, recall, and F1-score calculated with a macro-average approach to ensure fair multi-class performance assessment. The results show that AdaBoost and XGBoost achieve the highest test accuracy of 90.27%, while Random Forest and Gradient Boosting obtain 89.38%. Performance differences among models are relatively small (±1%) but indicate consistent predictive behavior. Feature importance analysis identifies BMI category and systolic blood pressure as the most influential predictors, followed by occupation and physical activity level, highlighting the relevance of lifestyle and physiological factors in sleep disorder risk. Overall, this study demonstrates that ensemble learning models provide reliable predictive performance and interpretable insights to support early detection of sleep disorders based on lifestyle patterns.

I. PENDAHULUAN

Kualitas tidur yang tidak baik merupakan masalah kesehatan global yang signifikan, terutama di kalangan mahasiswa, di mana kondisi ini terbukti berdampak langsung terhadap kemampuan konsentrasi, prestasi akademik, serta kesehatan mental dan fisik jangka panjang [1]. Tidur yang cukup memiliki peran vital dalam memperkuat fungsi kognitif, menjaga stabilitas emosi, dan mendukung kesehatan organ-organ penting seperti jantung, otak, serta sistem metabolisme pada berbagai kelompok usia [2]. Namun demikian, metode diagnosis konvensional seperti *polysomnografi (PSG)* memerlukan waktu pemeriksaan yang relatif lama, mencapai dua jam untuk satu siklus analisis tidur, serta bergantung pada ketersediaan fasilitas medis dan tenaga ahli yang terbatas. Kondisi ini menyebabkan proses deteksi dini gangguan tidur menjadi kurang efisien dan sulit diterapkan secara luas [3]. Berdasarkan data dari *National Institutes of Health*, sekitar sepertiga populasi dunia mengalami gangguan tidur yang berpotensi meningkatkan risiko depresi, obesitas, diabetes, serta penyakit kardiovaskular [4]. Oleh karena itu, diperlukan pendekatan diagnostik alternatif yang lebih cepat, efisien, dan mudah diimplementasikan untuk mendukung pengambilan keputusan klinis secara tepat.

Seiring dengan perkembangan teknologi, *machine learning* sebagai bagian dari kecerdasan buatan telah menjadi

pendekatan yang menjanjikan dalam bidang kesehatan, khususnya dalam analisis dan prediksi kondisi medis berbasis data. *Machine learning* memungkinkan pengembangan model komputasi yang mampu mempelajari pola kompleks dari data historis dan menghasilkan prediksi secara otomatis tanpa intervensi manusia secara langsung [5]. Dalam konteks gangguan tidur, pendekatan ini memiliki potensi besar untuk mengidentifikasi pola-pola tersembunyi yang sulit dideteksi melalui metode konvensional. Penelitian ini memanfaatkan dataset *Sleep Health and Lifestyle*, yang tidak hanya memuat data fisiologis, tetapi juga mencakup faktor gaya hidup yang berperan penting dalam kualitas tidur [6]. Proses klasifikasi digunakan untuk mengelompokkan individu ke dalam kategori gangguan tidur tertentu berdasarkan karakteristik dan fitur yang tersedia, sehingga diharapkan mampu menghasilkan model prediktif yang akurat dan relevan secara klinis [7]. Meskipun algoritma *machine learning* seperti *Logistic Regression* dan *Decision Tree* telah menunjukkan kinerja yang cukup baik dalam analisis data tidur, performa model tersebut masih dapat ditingkatkan, khususnya dalam hal akurasi dan kemampuan generalisasi, melalui penerapan metode *ensemble learning* yang mengombinasikan keunggulan beberapa algoritma sekaligus [8].

Ensemble learning merupakan pendekatan inovatif yang menggabungkan prediksi dari berbagai model dasar, seperti *averaging*, *bagging (Random Forest)*, *stacking*, serta *boosting (XGBoost* dan *AdaBoost)*, untuk membentuk model yang lebih robust dan stabil. Keunggulan utama metode ini terletak pada kemampuannya mengurangi risiko *overfitting* serta meningkatkan performa prediksi melalui keberagaman model penyusunnya [9]. Dalam konteks analisis gangguan tidur yang melibatkan data multidimensi dan berpotensi tidak seimbang, *ensemble learning* terbukti efektif dalam menangkap hubungan kompleks antarvariabel. Hal ini menjadikan *ensemble learning* sebagai salah satu pendekatan yang semakin banyak digunakan dalam penelitian kesehatan berbasis kecerdasan buatan dalam beberapa tahun terakhir [10].

Di sisi lain, untuk memaksimalkan kinerja model *ensemble learning*, proses penyetelan *hyperparameter* memegang peranan yang sangat krusial. Pemilihan kombinasi *hyperparameter* yang tidak optimal dapat menyebabkan penurunan performa model meskipun algoritma yang digunakan tergolong kuat [11]. Oleh karena itu, penelitian ini menerapkan *Grid Search Cross Validation* dengan skema *5-fold Cross Validation* pada data latih untuk memperoleh konfigurasi *hyperparameter* terbaik secara sistematis dan objektif. Sementara itu, data uji dipisahkan sepenuhnya untuk mengevaluasi performa akhir model secara tidak bias [12]. Pendekatan *k-fold Cross Validation* ini dinilai efektif dalam mengurangi risiko *overfitting*, terutama pada dataset berukuran terbatas yang belum sepenuhnya merepresentasikan populasi sebenarnya [13].

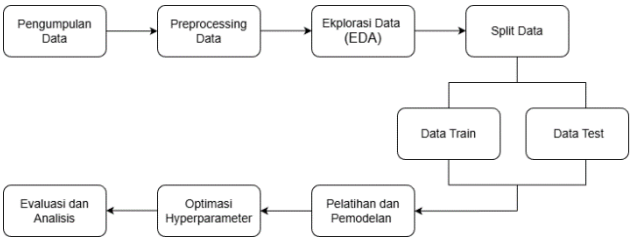
Berbagai penelitian sebelumnya telah mengevaluasi kinerja algoritma *ensemble learning* dalam klasifikasi gangguan tidur. Airlangga (2024) membandingkan *Random Forest* dan *XGBoost* dalam memprediksi gangguan tidur dan melaporkan akurasi tertinggi sebesar 92% untuk kedua algoritma [14]. Pardamean dkk. (2022) melakukan klasifikasi tahap tidur menggunakan *Random Forest* dan *XGBoost* berbasis data *wearable*, dengan *Random Forest* mencapai akurasi seimbang 78,5% dan *XGBoost* sebesar 79,5% [15]. Zhang dkk. (2024) melaporkan bahwa XGBoost menunjukkan performa terbaik dengan nilai *AUC* sebesar 87,2% pada studi terhadap 1.966 mahasiswa di Tiongkok [16]. Selain itu, penelitian oleh Ha dkk. (2023) dan Ramesh dkk. (2021) juga menegaskan keunggulan algoritma *ensemble* dalam memprediksi berbagai gangguan tidur seperti *obstructive sleep apnea* dan insomnia [17][18]. Meskipun hasil-hasil tersebut menunjukkan potensi yang menjanjikan, sebagian besar penelitian masih berfokus pada satu atau dua algoritma saja, serta menggunakan dataset yang terbatas pada aspek fisiologis atau data dari perangkat *wearable*.

Berbeda dari penelitian sebelumnya yang umumnya berfokus pada perbandingan nilai akurasi antar algoritma, penelitian ini menekankan stabilitas model melalui penerapan *Grid Search Cross Validation* dengan skema *5-fold*, evaluasi klasifikasi multi-kelas yang adil menggunakan pendekatan *macro-average*, serta interpretasi klinis faktor gaya hidup dan fisiologis melalui analisis *feature importance*. Pendekatan ini memberikan kontribusi tidak hanya secara prediktif, tetapi juga interpretatif, khususnya dalam konteks deteksi dini gangguan tidur berbasis pola gaya hidup.

Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk membandingkan performa empat algoritma *ensemble learning*, yaitu *Random Forest, Gradient Boosting, AdaBoost,* dan *XGBoost*, dalam memprediksi gangguan tidur berdasarkan pola gaya hidup menggunakan dataset *Sleep Health and Lifestyle*. Penelitian ini diharapkan mampu menghasilkan model prediksi yang lebih akurat dan efisien, sekaligus memberikan kontribusi ilmiah dalam pengembangan sistem pendukung keputusan berbasis machine learning di bidang kesehatan.

II. METODE

A. Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Penelitian ini diawali dengan *preprocessing* data untuk meningkatkan kualitas data sebelum pemodelan. Tahapan ini meliputi penanganan nilai kosong dengan pengisian label None, penghapusan atribut yang tidak relevan seperti Person

ID, standarisasi nilai kategori, serta transformasi atribut tekanan darah menjadi data numerik agar dapat diproses oleh algoritma pembelajaran mesin.

Selanjutnya, dilakukan analisis korelasi untuk memahami hubungan antar variabel dan mengidentifikasi pola yang berkontribusi terhadap gangguan tidur. Data kemudian dibagi menjadi data latih dan data uji untuk memastikan proses pelatihan dan pengujian dilakukan secara terpisah sehingga dapat mengurangi risiko *overfitting*.

Pada tahap pemodelan, digunakan beberapa algoritma ensemble learning, yaitu *Random Forest, Gradient Boosting, AdaBoost,* dan *XGBoost*, yang dipilih karena kemampuannya dalam meningkatkan akurasi dan stabilitas prediksi. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan *F1-score*.

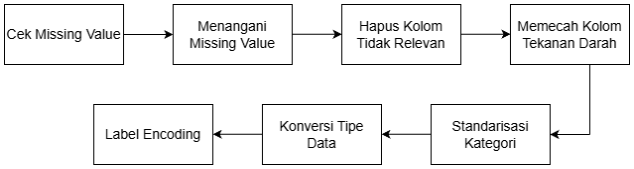
Dengan ini, untuk meningkatkan performa dan kemampuan generalisasi model, dilakukan optimasi hyperparameter menggunakan metode *Grid Search* dengan *5-fold Cross Validation*. Tahap akhir adalah evaluasi dan analisis model untuk menentukan algoritma terbaik dalam memprediksi gangguan tidur berdasarkan hasil evaluasi kinerja.

B. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah *Sleep Health and Lifestyle* yang diambil dari Kaggle, terdiri dari 374 sampel [19]. Meskipun ukuran dataset tergolong kecil, dataset ini tetap relevan karena mencakup kombinasi variabel fisiologis dan gaya hidup yang secara langsung berkaitan dengan gangguan tidur. Penggunaan teknik cross validation diterapkan untuk meminimalkan bias evaluasi akibat keterbatasan jumlah sampel. Dataset ini mencakup variabel seperti aktivitas fisik, tingkat stres, umur, jenis kelamin, profesi, tekanan darah, denyut jantung, langkah harian, lama tidur, kualitas tidur, dan keberadaan atau ketidakhadiran gangguan tidur termasuk dalam kumpulan data tersebut. Misalnya, kolom untuk gangguan tidur menunjukkan ‘Tidak Ada’ bagi mereka yang tidak mengalami gangguan tidur tertentu, ‘Insomnia’ untuk individu yang kesulitan untuk tidur atau mempertahankan tidur, dan *‘Sleep Apnea’* bagi individu yang mengalami gangguan pernapasan saat tidur [20].

C. Preprocessing Data

Teknik *preprocessing* merupakan langkah pertama dalam pengolahan data yang tabular, yang bertujuan untuk memperbaiki mutu data sebelum dievaluasi lebih dalam [21]. Tahapan preprocessing ini bertujuan untuk memastikan kualitas data yang optimal sehingga model ensemble learning dapat mempelajari pola secara efektif tanpa dipengaruhi oleh inkonsistensi data. Tahapan ini dilakukan agar model *machine learning* dapat beroperasi dengan lebih efisien menggunakan data yang lebih terstruktur untuk prediksi gangguan tidur. Alur proses *preprocessing* data ditunjukkan pada Gambar 2.



Gambar 2. Alur Preprocessing Data

1) Pengecekan Missing Value

Langkah pertama adalah pengecekan nilai yang hilang, data dievaluasi untuk menemukan nilai yang tidak terisi di setiap kolom. Sasaran dari proses ini adalah untuk memastikan tidak terdapat data yang hilang atau yang tidak lengkap [22].

2) Menangani Nilai Kosong

Kemudian, nilai kosong yang teridentifikasi diisi dengan label 'None' dengan arti tidak ada gangguan tidur untuk menjaga konsistensi data dan mencegah hilangnya informasi penting. Langkah ini sangat penting untuk memastikan dataset tetap lengkap, sehingga tidak ada kekurangan yang dapat mempengaruhi akurasi analisis akhir [23].

3) Menghapus Kolom Tidak Relevan

Pada tahap ini, penghapusan kolom-kolom yang tidak relevan, seperti *'Person ID'* untuk memastikan data hanya berisi informasi yang terkait dan sesuai dengan tujuan penelitian [24].

4) Pemisahan Kolom Tekanan Darah

Kolom tekanan darah dipisah menjadi *systolic_bp* dan *diastolic_bp* lalu dikonversi menjadi numerik sehingga mempermudah pengolahan data selama tahap preprocessing data [25].

5) Standarisasi Nilai Kategori

Berikutnya, dilakukan standarisasi terhadap nilai kategori untuk menyamakan format data, seperti mengubah label 'Normal Weight' menjadi 'Normal'. Langkah ini mengoptimalkan fitur dan meningkatkan karakteristik konvergensi algoritma, sehingga proses pelatihan model menjadi lebih efektif [26].

6) Konversi Tipe Data

Tahap selanjutnya adalah konversi tipe data untuk memastikan fitur numerik berada dalam format yang benar, seperti mengonversi kolom *systolic_bp* dan *diastolic_bp* dari string ke float. Langkah ini penting untuk menjamin bahwa semua atribut numerik memiliki format yang sesuai sebelum model dilatih [27].

7) Label Encoding

Tahap terakhir dari proses preprocessing adalah konversi variabel kategorikal ke bentuk numerik melalui teknik *Label Encoding*. Setiap label kategori dikonversi menjadi nilai numerik, misalnya ‘Insomnia’ menjadi 0, ‘None’ menjadi 1, dan ‘*Sleep Apnea*’ menjadi 2. Proses ini bertujuan untuk memastikan bahwa variabel kategorikal memiliki format

yang sesuai dengan kebutuhan model *machine learning*, sehingga algoritma dapat memahami, dan memanfaatkan fitur tersebut secara optimal selama proses pelatihan maupun pengujian model [28].

D. Eksplorasi Data

Setelah tahap *preprocessing*, langkah berikutnya adalah melakukan eksplorasi data untuk memahami karakteristik dan hubungan antar variabel pada dataset. Analisis dilakukan terhadap distribusi nilai setiap atribut, jumlah nilai unik pada masing-masing kolom, serta korelasi antar variabel numerik. Visualisasi *histogram* digunakan untuk menggambarkan distribusi atribut pada dataset. Analisis korelasi melalui *heatmap* membantu mengamati hubungan antara variabel gaya hidup, seperti tingkat stres, aktivitas fisik, dan durasi tidur, terhadap kualitas serta gangguan tidur. Tahap ini memberikan pemahaman awal mengenai pola data yang berpotensi memengaruhi hasil klasifikasi pada proses pemodelan [29].

E. Split Data

Dataset yang digunakan dalam penelitian ini terdiri dari 374 sampel, yang selanjutnya dibagi menjadi data pelatihan dan data pengujian dengan rasio 70% dan 30%. Proses pembagian data dilakukan menggunakan parameter *test_size = 0.3* dan *random_state = 42* untuk memastikan konsistensi hasil eksperimen. Dari total dataset, sekitar 261 sampel digunakan sebagai data pelatihan untuk melatih model ensemble learning agar mampu mempelajari pola kompleks secara optimal, sedangkan 113 sampel sisanya digunakan sebagai data pengujian untuk mengevaluasi kinerja akhir model secara independen [30].

Pemilihan rasio 70:30 didasarkan pada pertimbangan keseimbangan antara ketersediaan data yang memadai untuk proses pembelajaran model dan jumlah data uji yang representatif untuk evaluasi performa. Selain itu, untuk mengurangi potensi bias akibat penggunaan satu skema pembagian data, diterapkan *5-fold cross-validation* pada data pelatihan selama proses optimasi *hyperparameter*. Pendekatan ini bertujuan untuk meningkatkan stabilitas model serta memperbaiki kemampuan generalisasi, terutama mengingat ukuran dataset yang relatif terbatas.

F. Pelatihan dan Pemodelan

Pada tahap ini, dilakukan proses pelatihan dan pemodelan menggunakan algoritma *ensemble learning* yang meliputi *Random Forest, Gradient Boosting, AdaBoost,* dan *XGBoost*. Setiap model dilatih menggunakan data pelatihan (70%) melalui fungsi *fit (x_train, y_train)* untuk mempelajari pola hubungan antar variabel yang memengaruhi gangguan tidur. Selanjutnya, model yang telah dilatih digunakan untuk melakukan prediksi terhadap data pengujian (30%) menggunakan fungsi *predict (x_test)*.

Sebagai konteks pembanding, performa model *ensemble* yang diusulkan dalam penelitian ini dianalisis dengan mengacu pada hasil model *non-ensemble* yang telah dilaporkan pada penelitian sebelumnya, seperti *Logistic Regression* dan *Decision Tree*, guna menegaskan keunggulan pendekatan *ensemble learning* dalam meningkatkan stabilitas dan kinerja prediksi. Pendekatan ini digunakan untuk memberikan perspektif komparatif tanpa melakukan pelatihan ulang model *non-ensemble* secara langsung.

Tahap ini bertujuan untuk membandingkan performa keempat algoritma *ensemble learning* dalam mengklasifikasikan gangguan tidur berdasarkan pola gaya hidup secara sistematis dan objektif [31].

1) Random Forest

Random Forest adalah model *ensemble* berbasis pohon keputusan yang menggunakan *bootstrap sampling* dan *majority voting* untuk menggabungkan prediksi tiap pohon, termasuk metode *bagging* [32]. Model ini efisien, membutuhkan sampel relatif sedikit, dan memiliki akurasi klasifikasi tinggi, dengan penerapan efektif pada pemantauan lingkungan dan penggunaan lahan. Prediksi akhirnya dihitung sebagai voting mayoritas dari semua pohon [33].

$$f(x) = mode\big(h1(x), h2(x), \dots, hn(x)\big) \quad (1)$$

$f(x)$ adalah prediksi akhir untuk input x , sedangkan $h1(x)$ hingga $hn(x)$ merupakan prediksi masing-masing pohon. Majority voting menjamin model Random Forest stabil dan akurat.

2) Gradient Boosting

Gradient Boosting adalah metode *ensemble* yang membangun model secara bertahap dengan meminimalkan fungsi kehilangan melalui *gradien*. Setiap model lemah $hm(x)$ dilatih berurutan untuk memperbaiki residual kesalahan dari model sebelumnya [34]. Prediksi akhir diberikan oleh persamaan berikut.

$$f(x) = f0(x) + \sum_{m=1}^{M} \gamma_m \, \mathrm{h}_m(x) \quad (2)$$

Dimana $F0(x)$ adalah model awal, γm bobot model ke-m, dan $hm(x)$ model lemah pada iterasi ke-m. Pendekatan ini meningkatkan akurasi secara bertahap dengan mengurangi kesalahan kumulatif.

3) AdaBoost

AdaBoost adalah algoritma *ensemble* yang bertujuan mengubah beberapa pengklasifikasi lemah menjadi pengklasifikasi yang lebih kuat melalui proses iteratif. Pada setiap iterasi, sampel yang salah diklasifikasikan diberikan bobot lebih tinggi sehingga model berikutnya fokus memperbaiki kesalahan tersebut. Setiap model lemah $hm(x)$ diberi bobot , αm sesuai akurasinya, dan prediksi akhir dihitung sebagai penjumlahan berbobot [35].

$$h(x) = sign\left(\sum_{m=1}^{m} \alpha_m \, h_m \, (x)\right) \quad (3)$$

Pendekatan adaptif ini meningkatkan akurasi dan membuat *AdaBoost* efektif dalam tugas klasifikasi.

4) XGBoost

XGBoost adalah pengembangan dari *Gradient Boosting* yang meningkatkan efisiensi melalui regularisasi, pemangkasan pohon, dan paralelisme. Algoritma ini membangun model secara bertahap dengan meminimalkan fungsi kehilangan ℓ dan menambahkan penalti Ω untuk mengurangi *overfitting* [36].

$$L(F) = \sum_{i=1}^{n} \ell\left(y_i, F_{(xi)}\right) + \sum_{i=1}^{M} \Omega(F_m) \quad (4)$$

Keunggulan *XGBoost* meliputi skalabilitas tinggi, akurasi pada dataset besar, serta kemampuan mengekstrak fitur penting.

G. Optimasi Hyperparameter

Pada tahap ini dilakukan optimasi *hyperparameter* menggunakan metode *Grid Search CV* yang divalidasi dengan *5-fold Cross Validation* untuk setiap algoritma *ensemble*. Proses ini bertujuan mencari kombinasi parameter terbaik yang dapat meningkatkan performa prediktif dan kemampuan generalisasi model, sekaligus mengurangi risiko *overfitting* [37].

H. Evaluasi dan Analisis

Setelah proses pelatihan dan pemodelan menggunakan algoritma *Random Forest*, *Gradient Boosting*, *AdaBoost*, dan *XGBoost* selesai dilakukan, tahap selanjutnya adalah evaluasi untuk menilai performa model. Evaluasi ini dilakukan menggunakan *confusion matrix* yang terdiri dari *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Matriks ini memberikan gambaran menyeluruh mengenai kemampuan model dalam membedakan antara prediksi benar dan salah.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

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

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 - Score = 2x\frac{Precision\ x\ Recall}{Recall\ x\ Precision} \quad (8)$$

Akurasi menunjukkan proporsi prediksi yang benar terhadap keseluruhan data uji, sementara presisi mengukur ketepatan model dalam mengidentifikasi kelas positif. Selain itu, recall mengukur kemampuan model mengenali semua data yang benar-benar positif, dan *F1-Score* menilai keseimbangan antara presisi dan recall.

Oleh karena itu, permasalahan yang ditangani merupakan klasifikasi multi-kelas, perhitungan precision, recall, dan *F1-score* dilakukan menggunakan pendekatan *macro-average* untuk memberikan bobot yang setara pada setiap kelas. Selain itu, evaluasi tidak hanya difokuskan pada selisih nilai akurasi antar model, tetapi juga pada stabilitas performa serta interpretasi fitur penting yang berkontribusi terhadap prediksi gangguan tidur. Pendekatan ini digunakan untuk memperoleh pemahaman yang lebih komprehensif mengenai kinerja dan karakteristik masing-masing algoritma *ensemble learning*.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data penelitian ini diperoleh dari platform Kaggle melalui dataset *Sleep Health and Lifestyle* yang terdiri dari 374 sampel dan 13 atribut. Dataset ini dipilih karena memiliki struktur tabular yang sesuai untuk permasalahan klasifikasi serta memuat variabel gaya hidup dan indikator fisiologis yang secara klinis berkaitan dengan gangguan tidur. Selain itu, dataset telah melalui proses kurasi sehingga memiliki konsistensi dan kelengkapan data yang memadai untuk analisis machine learning.

Atribut yang digunakan mencakup jenis kelamin, usia, pekerjaan, durasi dan kualitas tidur, tingkat aktivitas fisik, tingkat stres, kategori BMI, tekanan darah, denyut jantung, dan jumlah langkah harian. Variabel target adalah Sleep Disorder yang terbagi ke dalam tiga kelas, yaitu Insomnia, None, dan Sleep Apnea, sehingga penelitian diformulasikan sebagai permasalahan klasifikasi multi-kelas.

Meskipun ukuran dataset relatif terbatas, setiap atribut memiliki keterkaitan langsung dengan faktor risiko gangguan tidur yang telah banyak dilaporkan dalam literatur. Untuk meningkatkan stabilitas dan generalisasi model, penelitian ini menerapkan validasi silang pada tahap pelatihan dan evaluasi. Dengan karakteristik tersebut, dataset ini dinilai representatif sebagai dasar pengembangan dan evaluasi model klasifikasi berbasis *ensemble learning*. Struktur lengkap dataset disajikan pada Tabel I.

TABEL I
DATASET SLEEP HEALTH AND LIFESTYLE

No	Atribut	Tipe Data
1	Person ID	Kategorikal (Unique Identifier)
2	Gender	Kategorikal
3	Age	Numerik
4	Occupation	Kategorikal
5	Sleep Duration	Numerik
6	Quality of Sleep	Numerik
7	Physical Activity Level	Numerik
8	Stress Level	Numerik
9	BMI Category	Kategorikal

10	Blood Pressure	Kategorikal
11	Heart Rate	Numerik
12	Daily Steps	Numerik
13	Sleep Disorder	Kategorikal (Target)

B. Preprocessing

Tahap *preprocessing* dilakukan untuk memastikan data berada dalam kondisi bersih, terstruktur, dan siap digunakan dalam proses pemodelan. Proses ini mencakup pemeriksaan *missing value*, penanganan nilai kosong, penghapusan atribut yang tidak relevan, pemisahan kolom tekanan darah menjadi *systolic_bp* dan *diastolic_bp*, standarisasi nilai kategori, konversi tipe data numerik, serta penerapan label encoding pada variabel kategorikal. Langkah-langkah ini dilakukan untuk meningkatkan konsistensi dan kualitas data sehingga model *machine learning* dapat dilatih secara optimal.

1) Pengecekan Missing Value

Pemeriksaan missing value dilakukan untuk mengidentifikasi kelengkapan data pada setiap atribut. Hasil analisis yang disajikan pada Tabel 2 menunjukkan bahwa seluruh atribut fitur tidak memiliki nilai kosong. Namun, terdapat 219 nilai hilang pada atribut Sleep Disorder yang merupakan variabel target.

Kondisi ini mengindikasikan bahwa sebagian responden tidak melaporkan adanya gangguan tidur tertentu, sehingga diperlukan penanganan khusus pada tahap preprocessing agar data tetap dapat digunakan secara optimal dalam proses pemodelan.

TABEL II
PENGECEKKAN MISSING VALUE

Atribut	Jumlah Nilai Hilang
Person ID	0
Gender	0
Age	0
Occupation	0
Sleep Duration	0
Quality of Sleep	0
Physical Activity Level	0
Stress Level	0
BMI Category	0
Blood Pressure	0
Heart Rate	0
Daily Steps	0
Sleep Disorder	219

2) Menangani Nilai Kosong

Penanganan nilai kosong dilakukan pada atribut Sleep Disorder dengan menerapkan teknik imputasi kategorikal menggunakan fungsi *fillna()* dari pustaka *Pandas*. Nilai NaN digantikan dengan label ‘None’, yang merepresentasikan responden tanpa gangguan tidur.

Pendekatan ini dipilih karena Sleep Disorder merupakan variabel kategorikal dan nilai kosong tidak menunjukkan kesalahan pencatatan, melainkan kondisi tidak adanya gangguan tidur. Setelah proses imputasi, seluruh data dinyatakan lengkap dan siap digunakan pada tahap preprocessing selanjutnya.

3) Menghapus Kolom Tidak Relevan

Pada tahap ini, dilakukan penghapusan kolom yang tidak berkontribusi terhadap proses klasifikasi. Kolom Person ID dihapus karena hanya berfungsi sebagai identitas unik responden dan tidak memiliki hubungan langsung dengan variabel target. Proses penghapusan dilakukan menggunakan fungsi *df.drop(['Person ID'], axis=1, inplace=True)* dari pustaka Pandas. Langkah ini bertujuan untuk meningkatkan efisiensi komputasi dan memastikan model hanya memproses atribut yang relevan terhadap prediksi gangguan tidur.

4) Pemisahan Kolom Tekanan Darah

Kolom Blood Pressure yang berisi nilai gabungan seperti ‘120/80’ dipisahkan menjadi dua kolom, *Systolic_bp* dan *Diastolic_bp*, menggunakan fungsi *str.split()* pada pustaka *Pandas*. Kedua atribut kemudian dikonversi ke tipe data numerik, sedangkan atribut asli dihapus agar tidak terjadi redundansi data. Langkah ini dilakukan untuk memperjelas makna fisiologis tekanan darah dan meningkatkan ketepatan analisis data. Selain itu, pemisahan kolom ini membantu model *ensemble learning* dalam memahami pengaruh masing-masing komponen tekanan darah secara terpisah terhadap kualitas tidur, sehingga diharapkan dapat meningkatkan akurasi prediksi dan performa model secara keseluruhan.

TABEL III
SETELAH PEMISAHAN KOLOM TEKANAN DARAH

No	Systolic_bp	Diastolic_bp
0	126	83
1	125	80
2	125	80
3	140	90
4	140	90

Tabel III menampilkan hasil pemisahan kolom Blood Pressure menjadi dua kolom terpisah, yaitu *systolic_bp* dan *diastolic_bp*. Setiap baris mewakili satu sampel peserta, dengan nilai tekanan darah sistolik dan diastolik yang telah dikonversi menjadi tipe data numerik. Sebagai contoh, sampel pertama memiliki tekanan darah 126/83 mmHg, yang dipisahkan menjadi *systolic_bp* = 126 dan *diastolic_bp* = 83.

5) Standarisasi Nilai Kategori

Tahap ini dilakukan untuk memastikan konsistensi data pada atribut kategorikal. Nilai yang memiliki variasi penulisan, seperti “Normal Weight” pada atribut BMI

Category, diseragamkan menjadi "Normal" menggunakan fungsi *replace()* pada pustaka *Pandas*. Standarisasi ini bertujuan untuk mencegah satu kategori direpresentasikan sebagai beberapa label berbeda, yang dapat menimbulkan bias pada proses pembelajaran model. Dengan demikian, setiap kategori diproses sebagai entitas tunggal oleh algoritma machine learning, sehingga meningkatkan kualitas fitur dan mempersiapkan data secara optimal untuk tahap encoding berikutnya.

6) Konversi Tipe Data

Tahap konversi tipe data dilakukan untuk memastikan setiap atribut memiliki tipe data yang sesuai dengan karakteristik variabel dan kebutuhan algoritma *machine learning*. Kolom *systolic_bp* dan *diastolic_bp*, yang sebelumnya dipisahkan dari Blood Pressure, dikonversi menjadi tipe numerik *(float)* menggunakan fungsi *astype()* pada *pandas*. Kolom numerik lainnya sudah memiliki tipe data yang sesuai, sehingga tidak memerlukan konversi tambahan.

Konversi ini penting karena algoritma ensemble learning seperti Random Forest, Gradient Boosting, AdaBoost, dan XGBoost mengandalkan operasi numerik dalam proses pemisahan node dan perhitungan loss function. Dengan memastikan seluruh atribut numerik berada dalam format yang benar, proses pelatihan model dapat berjalan secara stabil dan akurat. Dataset hasil konversi ini selanjutnya siap digunakan pada tahap encoding variabel kategorikal dan pelatihan model.

TABEL IV
MENGUBAH SYSTOLIC_BP DAN DIASTOLIC_BP MENJADI TIPE DATA NUMERIK

No	Systolic_bp	Diastolic_bp
0	126.0	83.0
1	125.0	80.0
2	125.0	80.0
3	140.0	90.0
4	140.0	90.0

Tabel IV menunjukkan konversi kolom *systolic_bp* dan *diastolic_bp* menjadi tipe numerik *(float)*. Setiap baris merepresentasikan satu sampel, misalnya sampel pertama memiliki tekanan darah 126/83 mmHg yang menjadi *systolic_bp* = 126.0 dan *diastolic_bp* = 83.0.

7) Label Encoding

Label Encoding dilakukan untuk mengubah kolom kategorikal menjadi format numerik agar dapat diproses oleh algoritma *machine learning*. Kolom yang di-encode meliputi Gender, Occupation, BMI Category, dan Sleep Disorder (target). Proses ini menggunakan LabelEncoder dari pustaka *Scikit-learn.*

Penggunaan Label Encoding dipilih karena algoritma ensemble learning berbasis pohon keputusan, seperti Random Forest, Gradient Boosting, AdaBoost, dan XGBoost, tidak bergantung pada hubungan ordinal antar nilai numerik, melainkan pada proses pemisahan (splitting) berdasarkan ambang batas (threshold). Dengan demikian, pemberian label numerik pada kategori tidak menimbulkan bias ordinal dalam proses pembelajaran model.

Langkah ini memastikan semua kategori diproses secara numerik, menjaga konsistensi data, dan mempersiapkan dataset untuk tahap pelatihan model *machine learning*.

TABEL V
HASIL LABEL ENCODING PADA ATRIBUT SLEEP DISORDER

Atribut	Kategori Asli	Nilai Encoding
Sleep Disorder	Insomnia	0
	None	1
	Sleep Apnea	2

Pada Tabel V menunjukkan bahwa hasil Label Encoding pada atribut *Sleep Disorder*, yang mengubah kategori teks menjadi angka untuk keperluan pemodelan. Kategori Insomnia menjadi 0, None menjadi 1, dan Sleep Apnea menjadi 2.

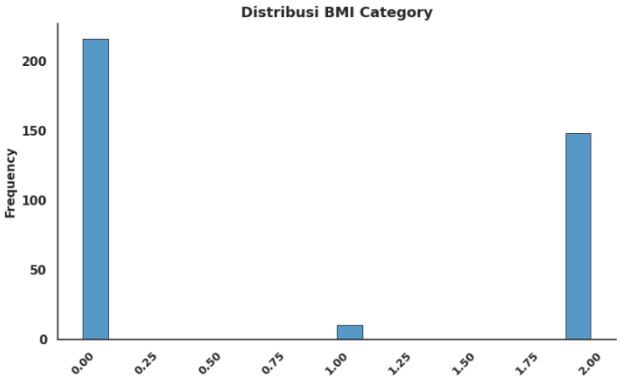
C. *Eksplorasi Data*

Setelah preprocessing, dataset dianalisis untuk memahami karakteristik, pola, distribusi nilai, hubungan antarvariabel, serta mendeteksi outlier atau anomali. Tahap ini penting untuk memberikan pemahaman awal terhadap data sehingga model *machine learning* dapat menghasilkan prediksi yang lebih akurat.

1) Analisis Distribusi Atribut Menggunakan Histogram

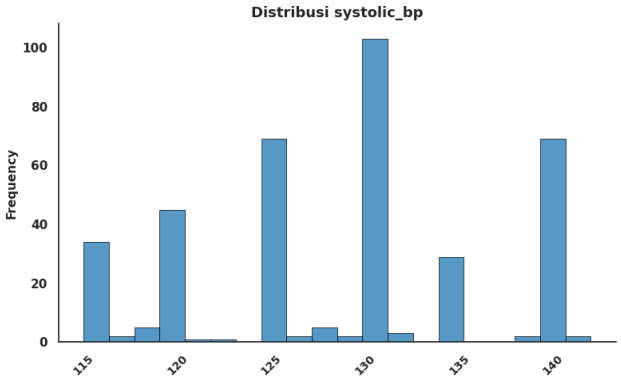
Untuk memberikan visualisasi yang jelas dan interpretasi klinis yang mendalam, analisis distribusi difokuskan pada empat atribut utama yang paling berpengaruh terhadap prediksi Sleep Disorder: BMI Category, Systolic_bp, Occupation, dan Physical Activity Level.

Distribusi BMI Category menunjukkan ketidakseimbangan kelas yang cukup jelas, di mana mayoritas responden berada pada kategori indeks massa tubuh normal dengan jumlah lebih dari 200 sampel. Sementara itu, kategori overweight dan obesitas mencakup sekitar 150 sampel, sedangkan kategori dengan proporsi lebih rendah relatif kurang terwakili. Kondisi ini menegaskan pentingnya penggunaan algoritma ensemble seperti XGBoost dan Random Forest yang dikenal lebih robust dalam menangani distribusi data yang tidak seimbang. Dari sisi klinis, temuan ini selaras dengan literatur yang menyatakan bahwa peningkatan BMI berkaitan erat dengan risiko terjadinya sleep apnea, sehingga memperkuat peran BMI sebagai fitur prediktif utama.



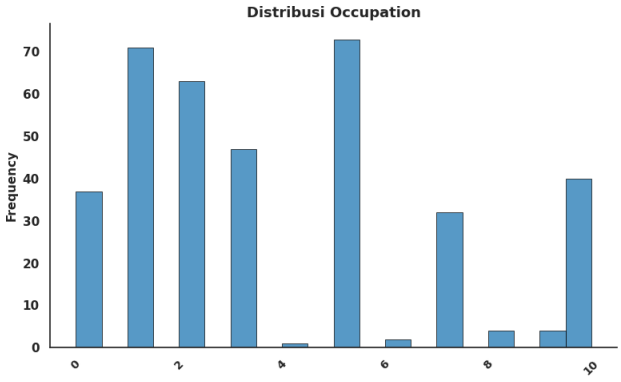
Gambar 3. Distribusi Atribut BMI Category

Pada atribut *Systolic_bp*, histogram menunjukkan pola distribusi multimodal dengan puncak frekuensi pada kisaran 125, 130, dan 140 mmHg. Sebagian besar responden berada dalam kategori pra-hipertensi hingga hipertensi tahap awal, yang secara klinis diketahui memiliki hubungan dengan gangguan tidur, khususnya *sleep apnea*. Pola distribusi ini menjelaskan mengapa tekanan darah sistolik muncul sebagai salah satu fitur paling berpengaruh dalam model prediksi.



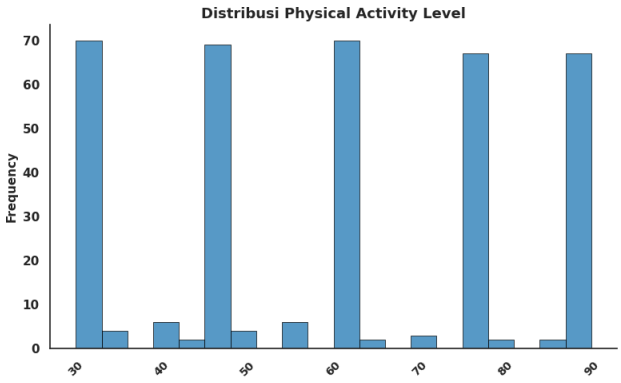
Gambar 4. Distribusi Atribut Systolic_bp

Distribusi atribut Occupation memperlihatkan keragaman profil subjek yang cukup luas, ditandai dengan dominasi beberapa kategori pekerjaan tertentu dan keberadaan kelas minoritas yang lebih kecil. Variasi jenis pekerjaan ini mencerminkan perbedaan tingkat stres, pola aktivitas harian, dan kebiasaan tidur, sehingga memberikan kontribusi penting sebagai faktor gaya hidup dalam proses klasifikasi gangguan tidur.



Gambar 5. Distribusi Atribut Occupation

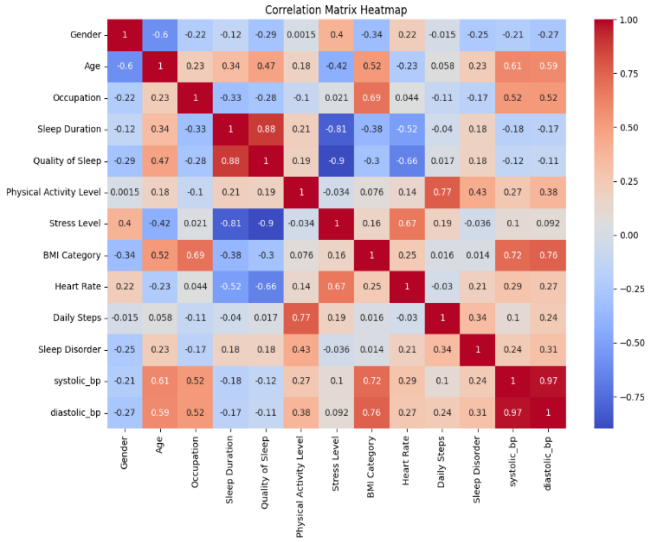
Sementara itu, atribut Physical Activity Level menunjukkan distribusi pada interval diskrit tertentu, seperti 30, 45, 60, 75, 80, dan 90 menit, dengan sebaran yang relatif merata. Pola ini memudahkan model dalam mengidentifikasi hubungan antara tingkat aktivitas fisik yang lebih tinggi dengan kualitas tidur yang lebih baik. Secara keseluruhan, analisis histogram ini tidak hanya mendukung hasil feature importance, tetapi juga memperkuat interpretasi klinis bahwa faktor fisiologis dan gaya hidup memiliki peran krusial dalam prediksi gangguan tidur.



Gambar 6. Distribusi Atribut Physical Activity Level

2) Peta Korelasi

Analisis korelasi menggunakan *Correlation Matrix Heatmap* difokuskan pada atribut yang relevan secara klinis: BMI Category, Systolic BP, Physical Activity Level, Sleep Duration, Quality of Sleep, dan Stress Level. Analisis ini juga memandu pemilihan fitur untuk evaluasi multi-kelas.



Gambar 7. Peta Korelasi

Korelasi antara fitur dan variabel target, yaitu Sleep Disorder, menunjukkan beberapa hubungan yang signifikan. Tingkat aktivitas fisik *(Physical Activity Level)* memiliki korelasi positif moderat dengan *Sleep Disorder (r = 0.43),* yang menunjukkan bahwa aktivitas fisik berkontribusi terhadap identifikasi jenis gangguan tidur. Jumlah langkah harian *(Daily Steps)* dan tekanan darah diastolik *(Diastolic_bp)* juga menunjukkan korelasi positif, masing-masing sebesar *(r = 0.34)* dan *(r = 0.31)*, yang mengindikasikan bahwa peningkatan tekanan darah diastolik dan aktivitas fisik terkait dengan variasi *Sleep Disorder*. Sementara itu, variabel gender memiliki korelasi negatif *(r = -0.25)*, yang mencerminkan adanya perbedaan pola gangguan tidur antara laki-laki dan perempuan.

Selain itu, hubungan antar faktor gaya hidup dan fisiologis menunjukkan korelasi yang kuat. BMI Category berkorelasi positif dengan tekanan darah sistolik *(r = 0.72)*, yang mengindikasikan bahwa BMI tinggi berkaitan dengan tekanan darah tinggi dan mendukung risiko Sleep Apnea. Tingkat aktivitas fisik dan jumlah langkah harian juga memiliki korelasi positif yang kuat *(r = 0.77)*, menandakan konsistensi data aktivitas fisik. Durasi tidur dan kualitas tidur menunjukkan korelasi yang sangat kuat *(r = 0.88)*, sehingga durasi tidur yang lebih lama cenderung berkaitan dengan kualitas tidur yang lebih baik.

Korelasi negatif yang signifikan juga diamati pada variabel stres. Tingkat stres (Stress Level) memiliki korelasi negatif yang sangat kuat dengan kualitas tidur *(r = -0.90)*, menunjukkan bahwa stres tinggi menurunkan kualitas tidur dan meningkatkan risiko Insomnia, serta korelasi negatif yang substansial dengan durasi tidur *(r = -0.81)*, yang menunjukkan bahwa stres tinggi juga mempersingkat durasi tidur.

Fokus pada atribut kunci ini memungkinkan interpretasi klinis yang lebih spesifik, di mana BMI tinggi dikaitkan dengan risiko Sleep Apnea, tingkat stres tinggi berisiko menyebabkan Insomnia, dan durasi tidur ekstrem berdampak pada kualitas tidur yang rendah. Evaluasi multi-kelas dilakukan menggunakan macro-average untuk metrik precision, recall, dan F1-score, sehingga setiap kelas Sleep Disorder (Insomnia, None, Sleep Apnea) diberi bobot yang sama, menjamin evaluasi performa model yang adil dan komprehensif. Hasil analisis ini memberikan justifikasi ilmiah bagi pemilihan algoritma ensemble learning, yang mampu menangkap interaksi non-linear antar fitur, sekaligus mendukung keputusan dalam tahap pemodelan machine learning selanjutnya.

D. Split Data

Setelah tahap eksplorasi data, dataset dibagi menjadi 70% data training (261 sampel) dan 30% data testing (113 sampel). Pembagian dilakukan menggunakan *stratified sampling* untuk menjaga proporsi kelas Sleep Disorder tetap seimbang di kedua subset, sehingga setiap kelas (Insomnia, None, Sleep Apnea) terwakili secara proporsional. Strategi ini mengurangi risiko bias evaluasi akibat distribusi kelas yang tidak seimbang dan mendukung analisis multi-kelas yang lebih akurat. Fitur (X) yang digunakan meliputi seluruh atribut numerik dan kategorikal yang telah di-encode, sedangkan target (y) adalah atribut Sleep Disorder. Dengan skema ini, model dapat belajar pola dari data training dan dievaluasi secara objektif pada data testing.

Strategi pembagian ini juga memastikan bahwa algoritma ensemble seperti *Random Forest, AdaBoost, Gradient Boosting*, dan *XGBoost* dapat belajar dari distribusi kelas yang representatif, memaksimalkan kemampuan model menangkap pola kompleks antar fitur, dan mempermudah interpretasi hasil terutama terkait fitur fisiologis (BMI, tekanan darah) dan gaya hidup (aktivitas fisik, stres, pekerjaan) yang berhubungan dengan risiko gangguan tidur. Selain itu, proporsi kelas yang seimbang memungkinkan perhitungan metrik evaluasi multi-kelas, seperti precision, recall, dan *F1-score*, dilakukan secara adil untuk setiap kelas.

E. Optimasi Hyperparameter

Optimasi *hyperparameter* dilakukan menggunakan *Grid Search Cross Validation (Grid Search CV)* dengan skema *5-fold Cross Validation* untuk memperoleh kombinasi parameter terbaik pada setiap model ensemble. Pendekatan ini menyeimbangkan bias dan variansi sehingga model memiliki performa optimal, stabil, dan mampu menangkap pola kompleks antarfitur, khususnya hubungan faktor fisiologis dan gaya hidup yang berkontribusi terhadap risiko sleep disorder. Validasi silang juga mengurangi ketergantungan pada satu pembagian data (70:30), penting mengingat ukuran dataset yang relatif kecil.

Parameter yang diuji pada masing-masing algoritma *ensemble* disajikan pada Tabel VI, disesuaikan dengan karakteristik algoritma dan ukuran dataset agar proses optimasi efisien tanpa meningkatkan risiko *overfitting*.

TABEL VI
OPTIMASI HYPERPARAMETER PADA SETIAP MODEL ENSEMBLE

Model	Parameter yang Dioptimasi	Nilai Parameter yang Diuji
Random Forest	n_estimators, max_depth, min_samples_split	[100, 200], [None,10,20], [2, 5]
Gradient Boosting	n_estimators, learning_rate, max_depth	[100, 200], [0.01, 0.1, 0.2], [3, 5, 7]
AdaBoost	n_estimators, learning_rate	[50, 100, 200], [0.01, 0.1, 1.0]
XGBoost	n_estimators, learning_rate, max_depth	[100, 200], [0.01, 0.1, 0.2], [3, 5, 7]

Hasil optimasi memberikan parameter terbaik yang digunakan pada evaluasi data uji (Tabel VII). AdaBoost dan XGBoost mencapai akurasi tertinggi 90.27%, sementara Random Forest dan Gradient Boosting sedikit lebih rendah 89.38% namun stabil. Selisih ±1% lebih mencerminkan keterbatasan dataset daripada perbedaan algoritmik.

TABEL VII
HASIL OPTIMASI HYPERPARAMETER DAN AKURASI MODEL

Model	Parameter Terbaik	Skor CV Terbaik	Akurasi Uji
Random Forest	n_estimators: 200, max_depth: None, min_samples_split: 5	0.9160	89.38%
Gradient Boosting	n_estimators: 200, learning_rate: 0.01, max_depth: 3	0.9083	89.38%
AdaBoost	n_estimators: 100, learning_rate: 0.1	0.8890	90.27%
XGBoost	n_estimators: 100, learning_rate: 0.01, max_depth: 5	0.9120	90.27%

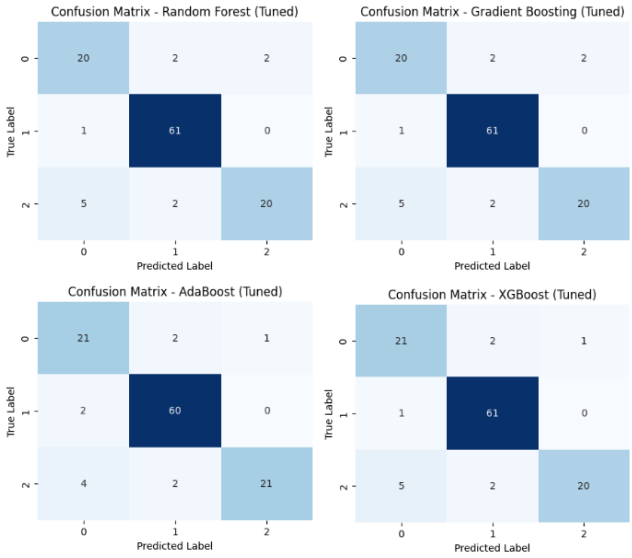
Evaluasi dilakukan pada klasifikasi multikelas (Insomnia, Sleep Apnea, None) menggunakan macro-average untuk precision, recall, dan F1-score. Pendekatan ini memberi bobot setara pada setiap kelas, menghindari bias kelas mayoritas, dan memudahkan interpretasi hasil, khususnya terkait fitur fisiologis dan gaya hidup.

5-fold Cross Validation memastikan stabilitas dan kemampuan generalisasi model, mengurangi risiko *overfitting*, terutama pada algoritma *boosting* yang sensitif terhadap *learning rate* dan jumlah estimator. Selain itu, model *ensemble* dapat dibandingkan dengan *baseline non-ensemble*, seperti *Logistic Regression* atau *SVM*, untuk menegaskan efektivitas metode ini dan membuka peluang pengembangan lebih lanjut.

Secara keseluruhan, strategi optimasi hyperparameter ini meningkatkan performa prediksi, stabilitas, evaluasi multikelas yang adil, dan interpretasi klinis yang relevan dalam konteks prediksi gangguan tidur berbasis gaya hidup.

F. Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan untuk menilai performa algoritma ensemble menggunakan hyperparameter terbaik hasil *Grid Search Cross Validation* dengan skema *5-fold Cross Validation*. Pendekatan ini memastikan bahwa evaluasi tidak bergantung pada satu pembagian data tertentu serta meningkatkan stabilitas dan kemampuan generalisasi model, khususnya mengingat ukuran dataset yang relatif terbatas. Evaluasi dilakukan pada klasifikasi multikelas yang terdiri dari tiga kelas, yaitu insomnia '0', none '1', dan sleep apnea '2'. *Confusion matrix* digunakan untuk memvisualisasikan hasil prediksi setiap model terhadap data uji, sebagaimana ditunjukkan pada Gambar 8.



Gambar 8. Visualisasi confusion matrix dari 4 model ensemble learning

Berdasarkan Gambar 8 tersebut, total sampel yang digunakan dalam proses evaluasi adalah 113 data untuk setiap model. Kinerja model diukur menggunakan metrik akurasi, yaitu rasio jumlah prediksi yang benar (True Positive) terhadap total sampel. Hasil visualisasi menunjukkan bahwa seluruh model memiliki pola confusion matrix yang relatif serupa, dengan perbedaan utama terdapat pada kemampuan mendeteksi kelas sleep apnea. Pada kelas ini, AdaBoost dan XGBoost mampu mengklasifikasikan lebih banyak sampel secara benar dibandingkan Random Forest dan Gradient Boosting. Jumlah *True Positive (TP)* masing-masing model tercantum pada Tabel VIII.

TABEL VIII
PERHITUNGAN AKURASI

Model	True Positive	Total Sampel	Akurasi
Random Forest	20+61+20=101	113	101/113=0.8938
Gradient Boosting	20+61+20=101	113	101/113=0.8938
AdaBoost	21+60+21=102	113	102/113=0.9027
XGBoost	21+61+20=102	113	102/113=0.9027

Berdasarkan Tabel VIII, terlihat bahwa *AdaBoost* dan *XGBoost* mampu mengklasifikasikan lebih banyak sampel dengan benar dibandingkan *Random Forest* dan *Gradient Boosting*, terutama pada kelas 2 (sleep apnea), yang menunjukkan keandalan kedua model dalam menangani kelas minoritas dengan relevansi klinis tinggi. Selain metrik akurasi, evaluasi kinerja model juga dilakukan menggunakan presisi, recall, dan F1-score, sehingga memberikan gambaran performa yang lebih komprehensif pada setiap kelas. Perbandingan hasil evaluasi antar model disajikan pada Tabel IX.

TABEL IX
PERBANDINGAN CONFUSION MATRIX MODEL ENSEMBLE LEARNING

Model	Akurasi	Presisi	Recall	F1-Score
XGBoost	90.27%	90.77%	90.27%	90.11%
AdaBoost	90.27%	90.76%	90.27%	90.23%
Gradient Boosting	89.38%	89.55%	89.38%	89.20%
Random Forest	89.38%	89.55%	89.38%	89.20%

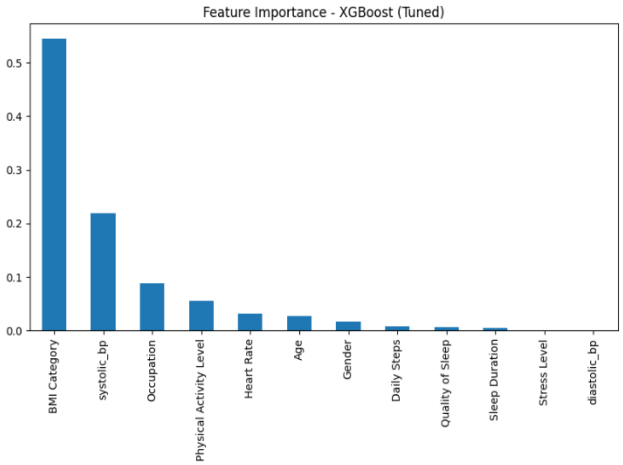
Metrik presisi, recall, dan F1-score dihitung menggunakan pendekatan macro-average, sehingga setiap kelas (insomnia, none, dan sleep apnea) memiliki bobot yang setara dalam proses evaluasi. Pendekatan ini penting untuk menghindari bias terhadap kelas mayoritas dan memastikan performa model pada kelas minoritas tetap terwakili secara adil.

Hasil yang ditunjukkan pada Tabel IX. Menunjukkan bahwa *XGBoost* dan *AdaBoost* memiliki performa terbaik setelah proses optimasi *hyperparameter*, dengan akurasi uji sebesar 90,27%. *XGBoost* menunjukkan nilai presisi tertinggi, sedangkan *AdaBoost* unggul pada nilai F1-score. Perbedaan performa antar model relatif kecil (±1%) dan belum diuji secara statistik, sehingga hasil ini lebih mencerminkan perbandingan empiris antar algoritma daripada klaim keunggulan signifikan.

Meskipun demikian, keunggulan model *ensemble* terlihat pada performa multikelas, terutama *AdaBoost* dalam mendeteksi kasus sleep apnea, yang menunjukkan kemampuan model dalam menangani kelas minoritas dengan relevansi klinis tinggi. Sebagai pembanding, model non-ensemble seperti *Logistic Regression* atau Support *Vector Machine (SVM)* dapat dipertimbangkan pada penelitian selanjutnya untuk menegaskan efektivitas pendekatan ensemble learning. Setelah model terbaik ditentukan, analisis feature importance dilakukan untuk mengevaluasi kontribusi masing-masing variabel terhadap hasil prediksi.

G. Analisis Pentingnya Fitur

Analisis pentingnya fitur dilakukan untuk mengidentifikasi kontribusi relatif setiap variabel terhadap hasil prediksi gangguan tidur. Analisis ini bertujuan tidak hanya untuk memahami perilaku model, tetapi juga untuk menafsirkan faktor-faktor fisiologis dan gaya hidup yang memiliki relevansi klinis terhadap risiko gangguan tidur. Meskipun *AdaBoost* memiliki F1-score sedikit lebih tinggi dan sensitif terhadap kelas minoritas seperti sleep apnea, *XGBoost* dipilih untuk analisis fitur guna memastikan kontribusi variabel yang dihitung mencerminkan pola yang stabil pada dataset yang relatif kecil.



Gambar 9 Visualisasi Pentingnya Fitur Dari Model XGBoost

Berdasarkan Gambar 9, fitur *BMI Category* memiliki kontribusi tertinggi dengan skor sekitar 0,55. Temuan ini menunjukkan bahwa indeks massa tubuh merupakan faktor dominan dalam prediksi gangguan tidur. Secara klinis, kondisi *overweight* dan obesitas telah banyak dilaporkan berkaitan erat dengan peningkatan risiko sleep apnea, akibat penumpukan jaringan lemak yang dapat menyempitkan saluran pernapasan dan mengganggu mekanisme pernapasan saat tidur.

Fitur *Systolic Blood Pressure* menempati posisi kedua dengan skor sekitar 0,22. Tekanan darah sistolik yang tinggi mencerminkan gangguan pada sistem kardiovaskular dan sering dikaitkan dengan kualitas tidur yang buruk serta gangguan tidur kronis. Hubungan dua arah antara hipertensi dan gangguan tidur, khususnya sleep apnea, telah dilaporkan secara luas dalam literatur medis, sehingga temuan ini memperkuat validitas klinis model yang dikembangkan.

Fitur *Occupation* memiliki kontribusi sekitar 0,09, menunjukkan bahwa jenis pekerjaan dan pola aktivitas harian berperan dalam menentukan kualitas dan keteraturan tidur. Pekerjaan dengan tingkat stres tinggi, jam kerja tidak teratur, atau aktivitas fisik rendah berpotensi meningkatkan risiko

gangguan tidur. Selanjutnya, *Physical Activity Level* (0,06) menegaskan bahwa aktivitas fisik yang memadai berkontribusi positif terhadap regulasi tidur, sejalan dengan temuan bahwa aktivitas fisik berperan dalam menjaga ritme sirkadian dan kualitas tidur.

Fitur *Heart Rate* (0,035) dan *Age* (0,03) memberikan pengaruh moderat terhadap prediksi model. Variasi denyut jantung dapat mencerminkan respons fisiologis terhadap stres atau kondisi kesehatan tertentu, sementara faktor usia berhubungan dengan perubahan pola tidur secara alami. Sebaliknya, fitur seperti *Gender, Daily Steps, Quality of Sleep, Sleep Duration, Stress Level, dan Diastolic Blood Pressure* menunjukkan nilai kepentingan yang relatif rendah, yang mengindikasikan bahwa pengaruhnya telah terwakili oleh fitur fisiologis dan gaya hidup yang lebih dominan.

Secara keseluruhan, hasil analisis ini menunjukkan bahwa kombinasi faktor fisiologis (BMI dan tekanan darah) serta faktor gaya hidup (aktivitas fisik dan pekerjaan) memainkan peran utama dalam prediksi gangguan tidur. Temuan ini menegaskan bahwa pendekatan berbasis machine learning tidak hanya efektif dalam klasifikasi gangguan tidur, tetapi juga mampu memberikan interpretasi yang bermakna secara klinis, khususnya dalam konteks deteksi dini dan pencegahan gangguan tidur berbasis pola gaya hidup.

H. Pembahasan Hasil Penelitian

Hasil penelitian menunjukkan bahwa pendekatan ensemble learning efektif untuk memprediksi gangguan tidur berdasarkan kombinasi faktor gaya hidup dan indikator fisiologis. Empat algoritma *ensemble* yang diuji, yaitu *Random Forest, Gradient Boosting, AdaBoost*, dan *XGBoost*, menunjukkan performa yang relatif konsisten dengan tingkat akurasi berkisar antara 89,38% hingga 90,27%. Hal ini mengindikasikan bahwa seluruh model mampu menangkap pola dasar pada data, meskipun perbedaan kinerja antar algoritma tergolong kecil.

Model *AdaBoost* dan *XGBoost* memperoleh akurasi tertinggi sebesar 90,27%, serta menunjukkan performa yang lebih baik dalam mendeteksi kelas sleep apnea, yang merupakan kelas minoritas dengan relevansi klinis tinggi. Keunggulan ini sejalan dengan karakteristik algoritma boosting yang secara iteratif memperbaiki kesalahan prediksi sebelumnya, sehingga lebih sensitif terhadap pola kompleks dan data yang tidak seimbang. Temuan ini konsisten dengan hasil penelitian sebelumnya yang melaporkan bahwa metode boosting cenderung unggul dibandingkan metode bagging pada permasalahan klasifikasi dengan struktur data kompleks.

Meskipun demikian, selisih akurasi antar model yang hanya sekitar ±1% menunjukkan bahwa perbedaan performa tersebut belum dapat diklaim signifikan secara statistik. Perbedaan kecil ini kemungkinan dipengaruhi oleh keterbatasan ukuran dataset yang hanya terdiri dari 374 sampel, sehingga variasi pembagian data dapat berdampak langsung terhadap hasil evaluasi. Oleh karena itu, hasil penelitian ini lebih tepat diinterpretasikan sebagai perbandingan empiris antar algoritma ensemble, bukan sebagai klaim keunggulan absolut suatu model.

Dengan ini, untuk mengurangi potensi bias evaluasi akibat keterbatasan data, penelitian ini menerapkan *Grid Search Cross Validation* dengan skema *5-fold Cross Validation* pada tahap optimasi *hyperparameter*. Pendekatan ini terbukti meningkatkan stabilitas model dan mengurangi risiko *overfitting*, khususnya pada algoritma *boosting* yang sensitif terhadap parameter seperti learning rate dan jumlah estimator. Selain itu, evaluasi dilakukan pada permasalahan klasifikasi multi-kelas (insomnia, none, dan sleep apnea) menggunakan metrik macro-average untuk precision, recall, dan F1-score, sehingga setiap kelas memperoleh bobot yang setara dan hasil evaluasi menjadi lebih adil.

Dari sisi interpretabilitas, analisis feature importance pada model *XGBoost* menunjukkan bahwa *BMI Category* dan *Systolic Blood Pressure* merupakan faktor paling berpengaruh dalam prediksi gangguan tidur, diikuti oleh *Occupation* dan *Physical Activity Level*. Temuan ini memiliki relevansi klinis yang kuat, karena obesitas dan hipertensi telah banyak dilaporkan sebagai faktor risiko utama sleep apnea, sementara aktivitas fisik dan jenis pekerjaan berkaitan dengan pola tidur, tingkat stres, dan ritme sirkadian. Dengan demikian, hasil penelitian ini tidak hanya bersifat prediktif, tetapi juga memberikan wawasan berbasis data mengenai faktor gaya hidup dan fisiologis yang berkontribusi terhadap gangguan tidur.

Secara keseluruhan, penelitian ini menunjukkan bahwa ensemble learning merupakan pendekatan yang andal untuk prediksi gangguan tidur berbasis gaya hidup, terutama dalam konteks klasifikasi multi-kelas dan deteksi kelas minoritas. Namun, mengingat keterbatasan ukuran dataset dan belum dilakukannya pengujian signifikansi statistik, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar, menambahkan model pembanding non-ensemble seperti *Logistic Regression* atau *Support Vector Machine (SVM)*, serta melakukan uji statistik untuk memastikan perbedaan kinerja antar model. Dengan pengembangan tersebut, model yang dihasilkan diharapkan dapat memberikan kontribusi yang lebih kuat dalam sistem pendukung keputusan kesehatan dan deteksi dini gangguan tidur.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan ensemble learning efektif untuk mengklasifikasikan gangguan tidur berbasis kombinasi faktor gaya hidup dan indikator fisiologis pada dataset *Sleep Health and Lifestyle* yang terdiri dari 374 sampel. Tahap preprocessing meliputi penanganan nilai kosong, penghapusan atribut tidak relevan, pemisahan tekanan darah menjadi variabel numerik, standarisasi kategori, serta label encoding berhasil meningkatkan kualitas dan konsistensi data sehingga siap digunakan dalam pemodelan.

Empat algoritma ensemble yang diuji, yaitu *Random Forest, Gradient Boosting, AdaBoost*, dan *XGBoost*, menunjukkan performa yang relatif konsisten dengan akurasi uji pada rentang 89,38%–90,27%. Optimasi *hyperparameter* menggunakan *Grid Search Cross Validation* dengan skema *5-fold Cross Validation* terbukti meningkatkan stabilitas dan kemampuan generalisasi model, khususnya pada algoritma boosting yang sensitif terhadap learning rate dan jumlah estimator. Evaluasi multikelas menggunakan metrik macro-average memastikan penilaian yang adil bagi setiap kelas (insomnia, none, dan sleep apnea).

AdaBoost dan *XGBoost* memperoleh akurasi tertinggi (90,27%) serta menunjukkan keunggulan dalam mendeteksi kelas sleep apnea, yang merupakan kelas minoritas dengan relevansi klinis tinggi. Namun, selisih kinerja antar model yang relatif kecil (±1%) belum diuji secara statistik, sehingga hasil ini lebih tepat dipahami sebagai perbandingan empiris antar algoritma, bukan klaim keunggulan signifikan.

Analisis *feature importance* mengindikasikan bahwa *BMI Category* dan *Systolic Blood Pressure* merupakan faktor paling berpengaruh, diikuti oleh *Occupation* dan *Physical Activity Level*. Temuan ini memiliki relevansi klinis yang kuat dan konsisten dengan literatur, di mana obesitas dan hipertensi berkaitan erat dengan risiko sleep apnea, sementara aktivitas fisik dan karakteristik pekerjaan memengaruhi kualitas serta pola tidur.

Secara keseluruhan, penelitian ini menegaskan bahwa algoritma *ensemble* khususnya *AdaBoost* dan *XGBoost* andal untuk prediksi gangguan tidur berbasis gaya hidup pada skenario data terbatas. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, menambahkan model pembanding *non-ensemble* misalnya *Logistic Regression* atau *SVM*, serta melakukan uji signifikansi statistik dan eksplorasi teknik *ensemble* lanjutan seperti stacking atau voting untuk memperkuat generalisasi dan validitas temuan.

DAFTAR PUSTAKA

[1] X. Liu *et al.*, "Poor sleep quality and its related risk factors among university students," *Ann. Palliat. Med.*, vol. 10, no. 4, pp. 4479–4485, 2021, doi: 10.21037/apm-21-472.

[2] Fahruzi et al., "Asesmen ECG-Apnea satu sadapan untuk peningkatan akurasi klasifikasi gangguan tidur berdasarkan AdaBoost (Single lead ECG-apnea recordings assessment for improved accuracy in classification of sleep disorder based on AdaBoost)," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, vol. 9, no. 2, pp. 196–204, 2020, doi: 10.22146/jnteti.v9i2.159

[3] Y. Wang, S. Ye, Z. Xu, Y. Chu, J. Zhang, and W. Yu, "Research on Sleep Staging Based on Support Vector Machine and Extreme Gradient Boosting Algorithm," *Nat. Sci. Sleep*, vol. 16, pp. 1827–1847, 2024, doi: 10.2147/NSS.S467111.

[4] A. S. Zamani *et al.*, "The prediction of sleep quality using wearable-assisted smart health monitoring systems based on statistical data," *J. King Saud Univ. - Sci.*, vol. 35, no. 9, Dec. 2023, doi: 10.1016/j.jksus.2023.102927.

[5] J. Kufel *et al.*, "What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine," *Diagnostics*, vol. 13, no. 15, 2023, doi: 10.3390/diagnostics13152582.

[6] M. A. Rahman *et al.*, "Improving Sleep Disorder Diagnosis Through Optimized Machine Learning Approaches," *IEEE Access*, vol. 13, pp. 20989–21004, 2025, doi: 10.1109/ACCESS.2025.3535535.

[7] S. Salmon, A. Azahari, and H. Ekawati, "Perbandingan Kinerja Algoritma K-Nearest Neighbor dan Algoritma Random Forest Untuk Klasifikasi Data Mining Pada Penyakit Gagal Ginjal," *Build. Informatics, Technol. Sci.*, vol. 6, no. 3, pp. 1943–1953, 2024, doi: 10.47065/bits.v6i3.6476.

[8] M. Fadhiel Alie and R. Rahmanda, "Model Prediksi Gangguan Tidur berdasarkan Beberapa Faktor menggunakan Machine Learning Prediction of Sleep Disorders based on Several Factors using Machine Learning," 2024. [Online]. Available: www.jurnal.unimed.ac.id

[9] R. Sudiyarno, A. Setyanto, and E. T. Luthfi, "Peningkatan Performa Pendeteksian Anomali Menggunakan Ensemble Learning dan Feature Selection," *Creat. Inf. Technol. J.*, vol. 7, no. 1, p. 1, 2021, doi: 10.24076/citec.2020v7i1.238.

[10] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 2, pp. 757–774, 2023, doi: 10.1016/j.jksuci.2023.01.014.

[11] J. A. Ilemobayo, O. Durodola, A. Ogungbire, and A. Osinuga, "Hyperparameter Tuning in Machine Learning : A Comprehensive Review," vol. 26, no. 6, pp. 388–395, 2024, doi: 10.9734/jerr/2024/v26i61188.

[12] E. Shahat, B. Data, D. El Shahat, A. Tolba, M. Abouhawwash, and M. A. Basset, "Machine learning and deep learning models based grid search cross validation for short - term solar irradiance forecasting," *J. Big Data*, 2024, doi: 10.1186/s40537-024-00991-w.

[13] P. Charilaou and R. Battat, "Machine learning models and over-fitting considerations," vol. 28, no. 5, pp. 605–607, 2022, doi: 10.3748/wjg.v28.i5.605.

[14] G. Airlangga, "Evaluating Machine Learning Models for Predicting Sleep Disorders in a Lifestyle and Health Data Context," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 1, pp. 51–57, 2024, doi: 10.33387/jiko.v7i1.7870.

[15] B. Pardamean, A. Budiarto, B. Mahesworo, A. A. Hidayat, and D. Sudigyo, "Sleep Stage Classification For Medical Purposes: Machine Learning Evaluation For Imbalanced Data," Research Square, Preprint, 2022, doi: 10.21203/rs.3.rs-1208553/v1.

[16] L. Zhang *et al.*, "Utilizing machine learning techniques to identify severe sleep disturbances in Chinese adolescents: an analysis of lifestyle, physical activity, and psychological factors," *Front. Psychiatry*, vol. 15, 2024, doi: 10.3389/fpsyt.2024.1447281.

[17] S. Ha *et al.*, "Predicting the Risk of Sleep Disorders Using a Machine Learning-Based Simple Questionnaire: Development and Validation Study," *J. Med. Internet Res.*, vol. 25, no. 1, 2023, doi: 10.2196/46520.

[18] J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, "Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning," *Healthc.*, vol. 9, no. 11, 2021, doi: 10.3390/healthcare9111450.

[19] Fakhruddin Fakhruddin and Sefrika Entas, "Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Prediksi Kualitas Tidur pada Kesehatan," *Vitam. J. ilmu Kesehat. Umum*, vol. 3, no. 4, pp. 217–234, Sep. 2025, doi: 10.61132/vitamin.v3i4.1773.

[20] M. Mostafa Monowar *et al.*, "Advanced sleep disorder detection using multi-layered ensemble learning and advanced data balancing techniques," *Front. Artif. Intell.*, vol. 7, 2024, doi: 10.3389/frai.2024.1506770.

[21] R. H. Saputra and R. R. Suryono, "Perbandingan Algoritma SVM , Random Forest , dan Naive Bayes Terhadap Kasus Scam di Media Sosial Twitter," vol. 7, no. 2, pp. 907–919, 2025, doi: 10.47065/bits.v7i2.7236.

[22] A. B. Mawardi, R. S. Pradini, and M. S. Haris, "Komparasi Algoritma Boosting Untuk Prediksi Gangguan Tidur," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.7281.

[23] A. Widianti and I. Pratama, "Penanganan Missing Values Dan Prediksi Data Timbunan Sampah Berbasis Machine Learning," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 2, pp. 242–251, 2024, doi: 10.36341/rabit.v9i2.4789.

[24] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1136–1144, 2023, doi:

10.47065/josh.v4i4.3803.
[25] D. D. N. Cahyo and A. Sunyoto, "Analisis Perbandingan Klasifikasi dalam Data Mining pada Prediksi Hujan dengan menggunakan Algoritma LSTM dan GRU," *J. Sains dan Inform.*, vol. 11, no. 1, pp. 40–49, 2025, doi: 10.34128/jsi.v11i1.1212.
[26] E. S. M. El-Kenawy, A. Ibrahim, A. A. Abdelhamid, N. Khodadadi, L. Abualigah, and M. M. Eid, "Predicting Sleep Disorders: Leveraging Sleep Health and Lifestyle Data with Dipper Throated Optimization Algorithm for Feature Selection and Logistic Regression for Classification," *Comput. J. Math. Stat. Sci.*, vol. 3, no. 2, pp. 341–358, 2024, doi: 10.21608/cjmss.2024.290167.1053.
[27] E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. M. F. S. El-Salhi, "The effect of preprocessing techniques, applied to numeric features, on classification algorithms' performance," *Data*, vol. 6, no. 2, pp. 1–23, 2021, doi: 10.3390/data6020011.
[28] V. Pranith Reddy, "Applying Machine Learning Algorithms for the Classification of Sleep Disorders," *Int. J. Sci. Res. Eng. Manag.*, vol. 09, no. 05, pp. 1–9, 2025, doi: 10.55041/ijsrem48664.
[29] A. Data, E. Eda, S. Peraih, and M. Olimpiade, "Exploratory Data Analysis (EDA): A Study of Olympic Medallist," vol. 11, no. 3 , pp. 578–587, 2022. [Online]. Available: https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/1857
[30] B. Vrigazova, "The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems," vol. 12, no. 1, pp. 228–242, 2021, doi: 10.2478/bsrj-2021-0015
[31] R. R. Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia," vol. 19, no. 2, pp. 302–311, 2020, doi: 10.30812/matrik.v19i2.688
[32] L. Yin, B. Li, P. Li, and R. Zhang, "Research on stock trend prediction method based on optimized random forest," *CAAI Trans. Intell. Technol.*, vol. 8, no. 1, pp. 274–284, 2023, doi: 10.1049/cit2.12067.
[33] Z. Wang, Z. Zhao, and C. Yin, "Fine Crop Classification Based on UAV Hyperspectral Images and Random Forest," *ISPRS Int. J. Geo-Information*, vol. 11, no. 4, 2022, doi: 10.3390/ijgi11040252.
[34] N. H. Setyawan and N. Wakhidah, "Analisis perbandingan metode logistic regression, random forest, gradient boosting untuk prediksi diabetes," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 10, no. 1, pp. 150–162, 2025, doi: 10.29100/jipi.v10i1.5743
[35] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. London, England: MIT Press, 2022. [Online]. Available: https://probml.github.io/pml-book/book1.html
[36] XGBoost Team, "XGBoost: A Scalable Tree Boosting System." Accessed: Nov. 08, 2025. [Online]. Available: https://xgboost.readthedocs.io/en/stable/
[37] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis," pp. 1–21, 2021, doi: 10.3390/informatics8040079