

**ANALISIS EFISIENSI TIDUR BERDASARKAN FAKTOR DEMOGRAFI DAN KEBIASAAN HARIAN DENGAN METODE *RANDOM FOREST REGRESSION***

**Neternomia Maria Ausiladora Da Costa\*, Anief Fauzan Rozi**

Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana  
Yogyakarta, Jl. Jembatan Merah No.84C Gejayan, Kabupaten Sleman, Daerah Istimewa  
Yogyakarta, Indonesia

\*email: [neternomiadacosta@gmail.com](mailto:neternomiadacosta@gmail.com)

**Received: 2025-05-18 Accepted: 2025-06-26 Published: 2025-06-30**

**Abstrak**

Efisiensi tidur berperan penting bagi kesehatan fisik dan mental, namun dapat menurun karena berbagai faktor demografi dan kebiasaan harian. Penelitian ini bertujuan menganalisis efisiensi tidur berdasarkan faktor-faktor tersebut menggunakan metode *Random Forest Regression*. Pendekatan kuantitatif diterapkan dengan dataset dari Kaggle melalui tahapan *data*, *feature selection*, dan pemodelan dengan 100 pohon keputusan. Untuk menilai performa secara menyeluruh, model dibandingkan dengan *Linear Regression* dan *Support Vector Regression* (SVR). Hasil penelitian menunjukkan bahwa *Random Forest Regression* memberikan performa yang sangat baik dengan nilai MAE sebesar 0,0327, MSE sebesar 0,0015, RMSE sebesar 0,0390, dan koefisien determinasi ( $R^2$ ) sebesar 0,91 yang lebih unggul dibanding dua model lainnya pada data pengujian. Model mampu mengikuti pola fluktuasi efisiensi tidur dan diterapkan dalam sistem prediksi interaktif yang menerima input seperti usia, waktu tidur, durasi tidur, jumlah terbangun, dan kebiasaan harian lainnya. Salah satu hasil prediksi menunjukkan nilai efisiensi tidur sebesar 0,528. Nilai ini merupakan output regresi dalam skala kontinu (0–1) dan tidak menghasilkan klasifikasi secara langsung. Kategorisasi kualitas tidur seperti “Sangat Baik”, “Baik”, dan “Kurang” dilakukan secara *post-hoc* (setelah hasil prediksi diperoleh) berdasarkan ambang batas tertentu. Kesimpulannya, faktor demografi dan kebiasaan harian memiliki pengaruh yang dapat diprediksi terhadap efisiensi tidur, dengan *Random Forest Regression* terbukti efektif untuk menganalisis hubungan kompleks ini. Penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain dan menambahkan variabel seperti tingkat stres dan faktor lingkungan untuk analisis yang lebih komprehensif.

**Kata kunci:** Efisiensi Tidur, *Random Forest Regression*, *Machine Learning*, Kebiasaan Harian, Faktor Demografi

**Abstract**

*Sleep efficiency plays a vital role in both physical and mental health, yet it can decline due to various demographic factors and daily habits. This study aims to analyze sleep efficiency using the Random Forest Regression method based on these influencing factors. A quantitative approach was conducted on a dataset obtained from Kaggle, which underwent preprocessing, feature selection, and modeling using 100 decision trees. To evaluate overall performance, the model was compared with Linear Regression and Support Vector Regression (SVR). The results indicated that Random Forest Regression achieved the best performance, with an MAE of 0.0327, MSE of 0.0015, RMSE of 0.0390, and a coefficient of determination ( $R^2$ ) of 0.91 on the test set, outperforming the other models. The model was also deployed in an interactive prediction system that accepts inputs such as age, bedtime, sleep duration, number of awakenings, and other habits. One prediction result generated a sleep efficiency score of 0.528. This value is a continuous output on a 0–1 scale and does not provide direct classification. The categorization of sleep quality, such as “Excellent,” “Good,” or “Poor,” was*

*applied post-hoc based on established threshold values. The study concludes that demographic and behavioral factors significantly influence sleep efficiency and that Random Forest Regression is effective in modeling these complex relationships. Future research should consider exploring alternative machine learning algorithms and integrating additional variables such as stress levels and environmental conditions for a more comprehensive assessment.*

**Keywords:** Sleep Efficiency, Random Forest Regression, Machine Learning, Daily Habits, Demographic Factors

**How to cite (in APA style):** Da Costa, N. M. A., & Rozi, A. F. (2025). Analisis efisiensi tidur berdasarkan faktor demografi dan kebiasaan harian dengan metode random forest regression. *Jurnal Pendidikan Informatika Dan Sains*, 14(1), 160–174. <https://doi.org/10.31571/saintek.v14i1.9291>

Copyright (c) 2025 Neternomia Maria Ausiladora Da Costa, Anief Fauzan Rozi  
DOI: 10.31571/saintek.v14i1.9291

## PENDAHULUAN

Bidang kesehatan berkembang pesat, dengan inovasi teknologi yang membantu masyarakat memantau kondisi tubuh secara *real-time*, seperti perangkat *wearable* dengan sensor yang melacak detak jantung, jumlah langkah, dan pola tidur, memberikan wawasan berharga tentang kesehatan secara keseluruhan (Driller et al., 2023). Tidur merupakan kegiatan yang menyita banyak waktu tetapi memiliki manfaat yang baik untuk kesehatan fisik dan mental. Kualitas tidur tidak hanya ditentukan oleh durasi tidur, tetapi dipengaruhi oleh berbagai faktor yang berkontribusi terhadap efisiensi tidur (Altini & Kinnunen, 2021). Efisiensi tidur (*sleep efficiency*) merupakan rasio total waktu tidur terhadap waktu di tempat tidur, yang dapat menurun secara signifikan seiring bertambahnya usia (Desjardins et al., 2019). Berdasarkan data *National Sleep Foundation* (January 7, 2025) menunjukkan bahwa kurang tidur berdampak negatif pada produktivitas kerja 60% orang Amerika. Lebih dari 2/3 (69%) orang Amerika melaporkan sulit berpikir jernih, dan enam dari sepuluh orang mengalami masalah dalam menangani beban kerja dan menghindari kesalahan (masing-masing 60%) jika mereka kurang tidur.

Penurunan efisiensi tidur berkaitan erat dengan penambahan usia. Pada kelompok usia 39–49 tahun, efisiensi tidur tercatat sebesar 85%, kemudian menurun menjadi 81% pada usia 60–69 tahun, dan mencapai 76% pada usia di atas 80 tahun (Cici Haryati et al., 2022). Pada faktor jenis kelamin juga efisiensi tidur memiliki nilai 77% pada pria berusia 60–70 tahun dan menurun menjadi hanya 46% pada pria berusia 85–105 tahun (Mazzotti et al., 2014). Sedangkan lansia wanita memiliki kualitas tidur subjektif yang kurang, namun sekitar 38,3% memiliki efisiensi tidur yang baik (>85%) yang diduga terkait dengan perubahan hormonal dan fisiologis yang terjadi selama menopause (Prasani et al., 2023).

Kebiasaan harian yang juga dapat mempengaruhi efisiensi tidur, seperti konsumsi alkohol, kafein, merokok, olahraga dan juga durasi tidur. Asupan kafein yang berlebihan, terutama sebelum tidur, diketahui dapat mengganggu pola tidur dengan memperpanjang latensi tidur dan menurunkan kualitas tidur (Meiranny & Chabibah, 2022). Durasi tidur yang kurang dari tujuh jam dapat menyebabkan gangguan *mood* dan fungsi kognitif, serta kualitas tidur yang baik sangat penting untuk konsentrasi belajar dan prestasi akademik (Wulandari & Pranata, 2024). *National Sleep Foundation* juga merekomendasikan menghindari nikotin dan alkohol sebelum tidur yang dapat berdampak besar pada kualitas tidur. Aktivitas fisik moderat, seperti olahraga ringan, dapat meningkatkan kualitas tidur dengan memperbaiki pola hidup sehat dan mengurangi stres, meskipun aktivitas fisik yang berlebihan dapat menyebabkan kesulitan tidur karena ketegangan.

Beberapa penelitian terdahulu menunjukkan efektivitas penggunaan metode *Random Forest Regression* dalam berbagai bidang. Pada penelitian yang ada (Fadillah & Fauzan, 2024) menemukan bahwa *Random Forest Regression* memberikan akurasi prediksi tertinggi dengan nilai RMSE terendah sebesar 7,90% dan MAE terendah sebesar 4,72% untuk prediksi batas kredit. Penelitian lain

(Patlisan, 2023) membuktikan bahwa model *Random Forest Regression* memiliki akurasi mencapai 94,87% untuk prediksi kuantitas pembelian barang setelah optimalisasi variabel. Ada yang menyimpulkan bahwa *Random Forest Regression* secara umum menghasilkan prediksi tingkat kemiskinan yang lebih akurat dibandingkan algoritma *XGBoost Regression* (Prastiyo & Febriandirza, 2024). Efektivitas *Random Forest Regression* dalam bidang kesehatan juga telah dikaji, (Saragih et al., 2024) menemukan bahwa algoritma ini memberikan prediksi paling akurat untuk prevalensi stunting di Sumatera Utara. (Özen, 2024) juga membuktikan performa tinggi *Random Forest* dalam memprediksi kasus dan kematian harian akibat COVID-19 dengan akurasi 92,30% dan  $R^2$  sebesar 0,9886, mengungguli LSTM dan ARIMA. Namun, dalam konteks tertentu seperti prediksi laju penguapan, studi oleh (Penalun et al., 2023) *Support Vector Regression* (SVR) juga menunjukkan performa yang kompetitif dengan *Random Forest Regression*, hal ini menunjukkan bahwa efektivitas suatu algoritma sangat bergantung pada karakteristik data dan permasalahan yang dihadapi.

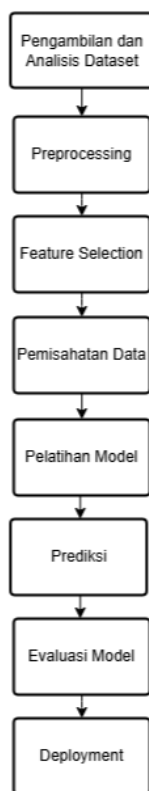
Selain itu, beberapa penelitian telah menyelidiki faktor-faktor yang memengaruhi efisiensi tidur. Melalui *Sleep Restriction Therapy* (SRT) menyatakan efisiensi tidur yang baik pada rentang 85-95%, kurang baik jika di bawah 85%, dan menyarankan penambahan waktu tidur jika efisiensi mencapai 90% atau lebih namun masih merasa kurang istirahat (Reed & Sacco, 2016). Penelitian (Halsen et al., 2022) menunjukkan adanya hubungan positif antara efisiensi tidur dan kualitas tidur, di mana tidur yang lebih awal dan konsisten dapat meningkatkan efisiensi tidur. Atlet dengan pola tidur teratur juga memiliki efisiensi tidur yang lebih tinggi serta variasi onset dan offset tidur yang lebih kecil. Sementara itu (Faoziyah & Suharjana, 2020), menemukan bahwa 60,3% mahasiswa dengan durasi tidur kurang dari tujuh jam mengalami gangguan konsentrasi dan penurunan semangat dalam pembelajaran gerak. Keteraturan jadwal tidur dan durasi tidur yang cukup juga terbukti berkorelasi positif dengan peningkatan efisiensi tidur dan performa pembelajaran motorik. Selain itu, hasil analisis regresi linier berganda oleh (He, 2023) menunjukkan bahwa aktivitas fisik dan konsumsi kalori memiliki pengaruh signifikan terhadap efisiensi tidur, peningkatan 1 km dalam jarak aktivitas sangat aktif dapat meningkatkan efisiensi tidur sebesar 0,012. Penelitian lebih lanjut mengonfirmasi bahwa efisiensi tidur berperan sebagai mediator antara tingkat aktivitas fisik dan fungsi kognitif, khususnya kontrol penghambatan (Li et al., 2021).

Penelitian ini mengembangkan studi sebelumnya yang umumnya membahas efisiensi tidur secara terbatas, baik dari sisi jumlah variabel maupun pendekatan analisis yang digunakan. Sebagian besar penelitian terdahulu hanya memfokuskan pada satu atau dua faktor, sehingga belum sepenuhnya merepresentasikan kondisi yang kompleks dalam kehidupan sehari-hari. Oleh karena itu, kebaruan dari penelitian ini terletak pada penggabungan berbagai faktor demografi dan kebiasaan harian dalam satu sistem prediktif yang menyeluruh, guna menganalisis efisiensi tidur secara lebih komprehensif. Metode *Random Forest Regression* adalah sebuah metode *data mining* yang merupakan algoritma *machine learning ensemble learning* digunakan untuk memprediksi nilai berdasarkan satu atau lebih variabel independen (Farhanuddin et al., 2024). Sistem yang dikembangkan memproses data seperti usia, jenis kelamin, waktu tidur, waktu bangun, durasi tidur, konsumsi alkohol dan kafein, status merokok, serta frekuensi olahraga. Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan pemodelan efisiensi tidur yang lebih akurat dan adaptif terhadap variasi data pengguna.

## **METODE**

### **Jalan Penelitian**

Jalan penelitian ini merupakan langkah-langkah yang digunakan untuk menghasilkan prediksi kualitas tidur dari nilai efisiensi tidur yang dianalisis. Adapun langkah-langkah yang harus dilakukan secara terstruktur dapat dilihat pada Gambar 1.



**Gambar 1. Diagram Alir Penelitian**

Adapun tahapan dalam alur penelitian sistem analisis pada penelitian ini:

1. Pengambilan Data

Penelitian ini menggunakan dataset dari *Kaggle.com* yang terdiri dari 16 atribut, yaitu *id*, *age*, *gender*, *bedtime*, *wakeup time*, *sleep duration*, *sleep efficiency*, *rem sleep percentage*, *deep sleep percentage*, *light sleep percentage*, *awakenings*, *caffeine consumption*, *alcohol consumption*, *smoking status*, dan *exercise frequency*. Dataset ini merepresentasikan faktor-faktor demografis, pola tidur, serta gaya hidup yang berkaitan dengan efisiensi tidur. Analisis awal dilakukan untuk memahami struktur data dan mendeteksi adanya nilai kosong maupun pola distribusi umum.

2. Preprocessing Data

Tahapan preprocessing mencakup transformasi kolom *bedtime* dan *wakeup time* ke format *datetime* serta pembuatan fitur *bedtime\_hour* dan *wakeup\_hour*. Nilai kosong pada beberapa atribut numerik ditangani dengan imputasi menggunakan nilai rata-rata. Selain itu, variabel kategorikal seperti *gender* dan *smoking status* dikodekan secara biner agar dapat digunakan dalam proses pemodelan.

3. Feature Selection

Seleksi fitur dilakukan dengan menghapus atribut yang tidak relevan seperti *id*, *bedtime*, dan *wakeup time*. Proses ini bertujuan untuk menyederhanakan model dan meningkatkan performa prediksi dengan mempertahankan fitur yang paling berpengaruh terhadap efisiensi tidur.

4. Pemisahan Data

Dataset dibagi menjadi 90% data pelatihan dan 10% data pengujian secara acak dengan parameter *random\_state = 101* yang tetap, untuk memastikan reproduksibilitas dan evaluasi yang objektif terhadap performa model.

## 5. Pemodelan Regresi

Metode *Random Forest Regression* digunakan sebagai model utama dengan parameter  $n\_estimators = 100$  dan  $max\_depth = 5$ . Model dilatih menggunakan data pelatihan untuk mempelajari hubungan antara variabel input dengan target prediksi, yaitu efisiensi tidur.

## 6. Prediksi Efisiensi Tidur

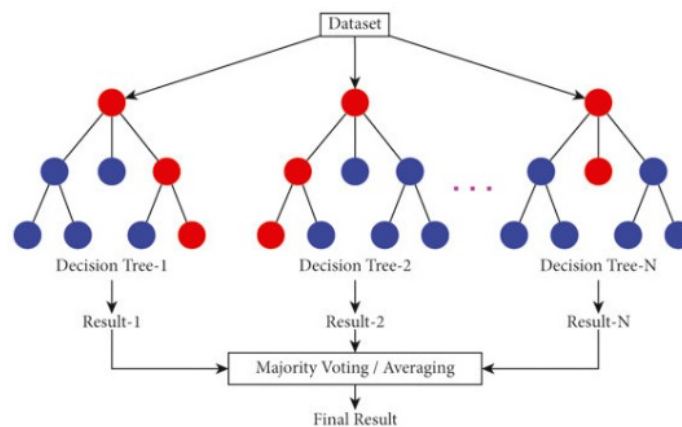
Model yang telah dilatih digunakan untuk memprediksi nilai efisiensi tidur pada data pengujian. Hasil prediksi dibandingkan dengan nilai aktual dan divisualisasikan untuk mengevaluasi kesesuaian pola dan tingkat akurasi prediksi.

## 7. Evaluasi Model

Evaluasi performa dilakukan menggunakan empat metrik regresi: MAE, MSE, RMSE, dan  $R^2$ . Sebagai pembanding, digunakan dua algoritma regresi lainnya yaitu *Linear Regression* dan *Support Vector Regression (SVR)* untuk menilai keunggulan model utama secara kuantitatif. Jika semua proses evaluasi sesuai dan berhasil mendapatkan hasil error yang kecil sehingga dapat memperkecil risiko dalam prediksi, maka proses yang selanjutnya yang dilakukan adalah *deployment*.

### Metode *Random Forest Regression*

*Random Forest* merupakan suatu metode *machine learning* yang digunakan untuk permasalahan klasifikasi dan regresi. Dan merupakan salah satu algoritma yang digunakan untuk meningkatkan akurasi, serta memiliki tingkat *error* yang cukup rendah dibandingkan *algoritma classifier* yang lainnya (Patlisan, 2023). *Random Forest* bermanfaat untuk mengurangi *overfitting* karena menggunakan banyak pohon keputusan (Prastiyo & Febriandirza, 2024). Algoritma ini meliputi sebagian besar pohon keputusan yang dilatih dan hasil akhirnya akan diambil rata-rata untuk bagian permasalahan regresi (Ricky Verdiyanto, Dwi Hartanti, 2025). Saat membangun satu pohon dalam *Random Forest*, pengacakan dilakukan untuk memilih *node* terbaik. Dalam permasalahan regresi, proses akan dimulai dengan *node root* dan kemudian dipisahkan berdasarkan hasil variabel sampai *leaf node* tercapai dan hasilnya diberikan (Patlisan, 2023). Contoh pohon keputusan dengan pola regresi dapat dilihat pada Gambar 2.



**Gambar 2. Contoh Pohon Keputusan Random Forest Regression**

Adapun proses yang dilalui dalam proses berjalannya algoritma *Random Forest Regression* dengan rumus-rumus yang digunakan untuk melakukan prediksi akhir sebagai berikut:

### a. Pembentukan Pohon

Dengan menggunakan teknik pengambilan sampel dengan penggantian, *Random Forest* membuat beberapa pohon keputusan dari data latih. Setiap pohon menggunakan subset fitur acak, yang menurunkan *overfitting* dan meningkatkan generalisasi (Hafizh et al., 2025).

### b. Prediksi

*Random Forest* menggabungkan hasil prediksi dari banyak pohon keputusan untuk mendapatkan prediksi akhir. Untuk regresi, hasil prediksi rata-rata dari semua pohon digunakan sebagai hasil akhir (Hafizh et al., 2025). Rumus untuk menghitung nilai rata-rata seluruh prediksi pohon adalah:

$$\hat{Y}_l = \frac{1}{N_{tree}} \sum_{n=1}^{N_{tree}} \hat{Y}_n$$

Keterangan:

$\hat{Y}_l$  = Hasil Prediksi akhir

$N_{tree}$  = Total Jumlah pohon pada Random Forest

$\hat{Y}_n$  = Hasil prediksi pohon ke-n

c. Evaluasi Model

Pengujian akurasi model dilakukan untuk memahami kinerja model dalam konteks data yang ada dan menentukan sejauh mana model dapat memberikan prediksi yang tepat. Dalam penelitian ini, metrik MSE, MAE, dan RMSE digunakan untuk mengukur akurasi model, yang menunjukkan tingkat kesalahan prediksi. Semakin rendah nilai error dari ketiga metrik, semakin baik akurasi prediksinya (Nur et al., 2024). Selain itu, metrik koefisien determinasi ( $R^2$ ) digunakan untuk mengevaluasi seberapa jauh kemampuan model untuk menjelaskan variasi pada variabel dependen (target). Nilai  $R^2$  berada dalam rentang 0 sampai 1, di mana nilai yang mendekati 1 menunjukkan bahwa model mampu menerangkan proporsi variasi target secara lebih baik (Sehangunaung et al., 2023).

Untuk mengukur varian residu, *Mean Squared Error* (MSE) adalah rata-rata dari perbedaan kuadrat antara nilai aktual dan nilai prediksi pada set data (Nugraha et al., 2023). Rumus MSE disajikan pada Gambar 1.

$$MSE = \sum \frac{(y_t - \hat{y}_t)^2}{n} \quad (1)$$

*Mean Absolute Error* (MAE) mengukur rata-rata residu pada suatu set data. Ini adalah representasi rata-rata dari perbedaan absolut antara nilai aktual dan nilai prediksi (Ardesfira et al., 2022). Rumus MAE disajikan pada Gambar 2.

$$MAE = \sum \frac{|y_t - \hat{y}_t|}{n} \quad (2)$$

Akar pangkat dari MSE adalah *Root Mean Squared Error* (RMSE). RMSE mengukur standar deviasi residu (Tjahjadi et al., 2022). Rumus RMSE disajikan Gambar 3

$$RMSE = \sqrt{\sum \frac{(y_t - \hat{y}_t)^2}{n}} \quad (3)$$

Sementara itu, koefisien determinasi ( $R^2$ ) nilai yang memperlihatkan seberapa besar variabel bebas memengaruhi variabel terikat (Ardiansyah, 2023). Rumus  $R^2$  disajikan pada Gambar 4.

$$R^2 = 1 - \frac{\sum (y_t - \hat{y}_t)^2}{\sum (y_t - \bar{y})^2} \quad (4)$$

Keterangan:

- $Y_t$  = Nilai prediksi dari model regresi  
 $\hat{Y}_t$  = Nilai aktual dari variabel dependen  
 $\bar{Y}$  = Rata-rata nilai aktual dari variabel dependen  
 $n$  = Jumlah total observasi atau data

## HASIL DAN PEMBAHASAN

Pada bagian ini berisi hasil dan pembahasan dari topik penelitian. Data yang dimiliki dilakukan *preprocessing* sebagai awal persiapan untuk pemrosesan *machine learning*.

### *Preprocessing*

Tahap *preprocessing* adalah langkah yang cukup penting dalam proses pemahaman dan analisis data yang berguna untuk memastikan kualitas data yang optimal sebelum proses pemodelan terjadi. Pada penelitian ini, *preprocessing* dilakukan melalui beberapa tahap, seperti melakukan proses pemeriksaan awal struktur data dengan mengamati tipe data tiap atribut dan mendeteksi keberadaan nilai kosong (*missing values*) untuk menentukan proses penanganan yang sesuai. Setelah pemeriksaan awal selesai, maka masuk dalam proses pengelolaan yang diawali dengan transformasi tipe data yang dimiliki kolom *bedtime* dan *wake up time* yang semula bertipe string dikonversi atau diganti menjadi tipe data *datetime*. Proses ini dilakukan untuk memudahkan proses pengambilan data terkait waktu untuk melakukan modifikasi data menjadi fitur baru yaitu *bedtime\_hour* dan *wakeup\_hour*. Fitur baru ini nilainya didapatkan dari proses pengambilan komponen jam dan menit dari data waktu tidur dan bangun dikumpulkan, dan kemudian diubah menjadi format jam desimal, dengan tujuan untuk mempermudah analisis terkait waktu yang berkaitan dengan kualitas tidur.

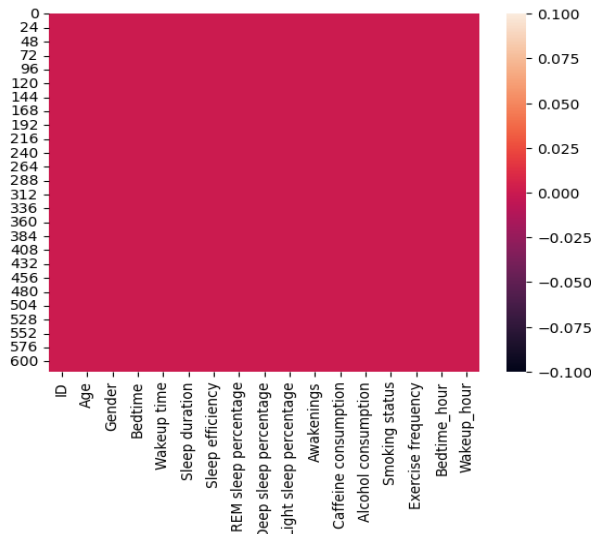
Selanjutnya diketahui juga pada data terdapat *missing values* pada variabel *numerik* seperti *awakenings*, *caffeine consumption*, *alcohol consumption*, dan *exercise frequency* yang ditunjukkan pada Tabel 1.

**Tabel 1. Jumlah *Missing Values* Tiap Variabel**

| Nama Variabel          | Jumlah <i>Missing Values</i> |
|------------------------|------------------------------|
| ID                     | 0                            |
| Age                    | 0                            |
| Gender                 | 0                            |
| Bedtime                | 0                            |
| Wakeup time            | 0                            |
| Sleep duration         | 0                            |
| Sleep efficiency       | 0                            |
| REM sleep percentage   | 0                            |
| Deep sleep percentage  | 0                            |
| Light sleep percentage | 0                            |
| Awakenings             | 27                           |
| Caffeine Consumption   | 36                           |
| Alcohol Consumption    | 21                           |
| Smoking status         | 0                            |
| Exercise frequency     | 7                            |

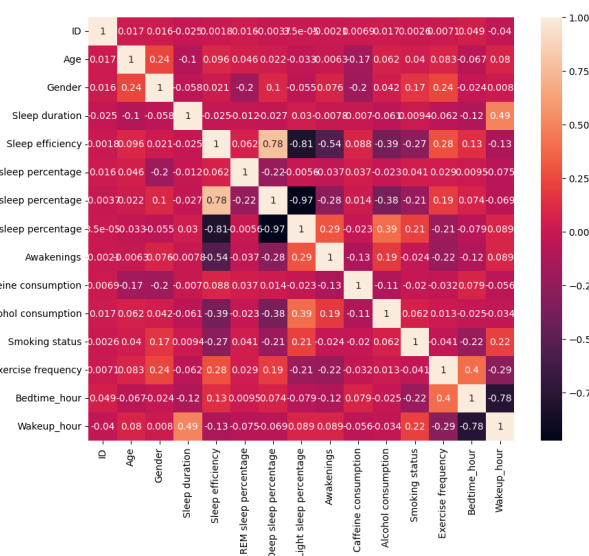
Pada Tabel 1 dapat terlihat pada kolom *awakenings* memiliki 27 *missing values*, pada kolom *caffeine consumption* 36 *missing values*, kolom *alcohol consumption* 21 *missing values*, dan kolom *exercise frequency* memiliki 7 *missing values*. Data-data yang hilang (*missing values*) pada variabel tersebut dilengkapi dengan menggunakan teknik imputasi rata-rata, maksud dari proses ini yaitu data-data yang hilang ini diisi dengan rata-rata perhitungan dari isi data tiap kolom variabel. Metode ini digunakan karena kemampuannya yang dapat menjaga pola distribusi data asli serta bisa

meminimalkan kemungkinan munculnya bias sistematis atau penyimpangan dalam analisis. Setelah proses imputasi selesai, dilakukan pengecekan ulang terhadap kondisi dataset agar memastikan tidak lagi ada *missing values* yang divisualisasikan menggunakan *heatmap*, yang ditunjukkan pada Gambar 3.



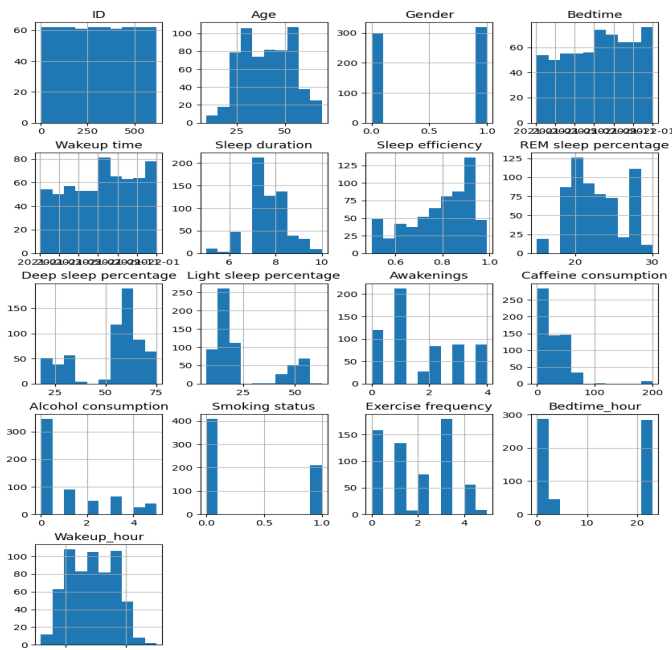
Gambar 3. *Heatmap Missing Values* pada Dataset

Proses lainnya yaitu memproses data pada kolom yang masih bertipe data *object* untuk dilakukan transformasi variabel kategorikal atau singkatnya sebagai *encoding*. proses *encoding* ini diterapkan agar nantinya dapat digunakan untuk pemodelan pada dua atribut utama dalam dataset, yaitu pada variabel *gender* dan *smoking status* untuk konversi yang dilakukan pada data tiap kolom variabel dengan merepresentasikan biner menggunakan nilai 1 dan 0. adapun transformasinya untuk variabel *gender* menetapkan nilai 1 untuk kategori “*male*” dan nilai 0 untuk kategori “*female*”, sedangkan untuk variabel *smoking status* mengalokasikan nilai 1 untuk responden yang menjawab “*yes*” dan 0 untuk jawaban “*no*”. Untuk memberikan pemahaman tentang hubungan antara variabel numerik dalam dataset, dilakukan pembuatan visualisasi heatmap korelasi menggunakan metode koefisien korelasi pearson, visualisasi ini menyajikan pandangan mengenai bagaimana variabel-variabel saling berhubungan, dengan fokus khusus pada interaksi antara berbagai fitur input dengan variabel target yaitu *sleep efficiency*. Hasil analisis korelasi ini dapat dilihat pada Gambar 4.



Gambar 4. *Heatmap Korelasi Variabel Numerik*

Hasil analisis korelasi menunjukkan bahwa *deep sleep percentage* memiliki korelasi positif paling kuat terhadap efisiensi tidur, diikuti oleh *exercise frequency*, sedangkan *awakenings* menunjukkan korelasi negatif. Pada bagian variabel demografi seperti usia, jenis kelamin, dan status merokok tidak menunjukkan korelasi yang signifikan, tetapi variabel ini tetap digunakan dalam pemodelan untuk menunjukkan hubungan nonlinear yang mungkin. Selain itu, dibuat histogram pada masing-masing fitur numerik untuk menganalisis karakteristik distribusi data. Histogram ini membantu dalam menilai sebaran data, menemukan *skewness*, dan menemukan potensi *outlier*. Gambar 5 menunjukkan histogram yang menunjukkan distribusi fitur-fitur ini.



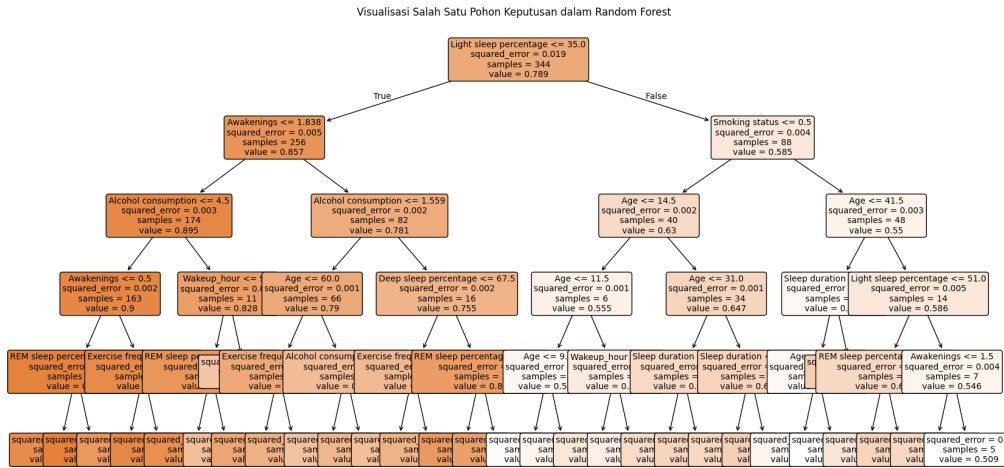
**Gambar 5. Histogram Distribusi Fitur Numerik**

Gambar 5 menunjukkan bahwa beberapa variabel memiliki distribusi yang tidak simetris atau *skewness*. Misalnya, variabel *caffeine consumption* dan *alcohol consumption* menunjukkan distribusi yang miring ke kanan, yang berarti sebagian besar responden memiliki konsumsi yang rendah. Namun, beberapa orang menunjukkan konsumsi yang sangat tinggi. Selain itu, ada *skewness* ke kanan pada variabel *sleep duration*, yang menunjukkan bahwa mayoritas responden memiliki *sleep duration* yang relatif pendek. Sebaliknya, variabel seperti *sleep efficiency*, *deep sleep percentage*, dan *light sleep percentage* menunjukkan distribusi yang lebih merata. Sebaliknya, batang diskrit digunakan untuk menampilkan atribut kategorikal seperti *gender*, *smoking status*, dan *exercise frequency*. Selain itu, pola tidur responden dapat digambarkan dengan membagi waktu tidur (*bedtime\_hour*) dan waktu bangun (*wakeup\_hour*) dengan mayoritas waktu tidur berada pada malam hari dan waktu bangun antara pukul 5 hingga 8 pagi. Setelah menyelesaikan seluruh tahapan *preprocessing*, dataset yang semula memiliki berbagai kekurangan telah berhasil ditransformasi menjadi bentuk yang lebih bersih, konsisten, dan optimal untuk digunakan dalam proses pemodelan selanjutnya.

### **Pemodelan *Random Forest Regression***

Setelah tahap *preprocessing*, proses pemodelan dimulai dengan *feature selection* untuk meningkatkan kinerja model dan mengurangi kompleksitas data. Dari 16 atribut awal, tiga fitur dihapus: *bedtime* dan *wakeup time* karena telah direpresentasikan oleh *bedtime\_hour* dan *wakeup\_hour*, serta *ID* karena bersifat unik dan tidak berkontribusi pada prediksi. Selain itu, *sleep efficiency* ditetapkan sebagai target (*y*), sehingga tidak dimasukkan sebagai fitur input. hasilnya, terdapat 13 fitur yang digunakan sebagai input (*X*), mencakup faktor demografi dan kebiasaan harian. Dataset kemudian dibagi menjadi 90% data latih dan 10% data uji dengan *random\_state=101* untuk

memastikan replikasi yang konsisten. Model *Random Forest Regression* dilatih dengan parameter  $n\_estimators=100$  dan  $max\_depth=5$  untuk menjaga keseimbangan antara akurasi dan generalisasi. Pelatihan dilakukan pada data latih ( $X\_train, y\_train$ ) dengan pohon-pohon keputusan yang dibentuk dari subset acak data dan fitur. Salah satu pohon divisualisasikan untuk menggambarkan struktur model dataset ini, seperti ditampilkan pada Gambar 6.



**Gambar 6. Pohon Keputusan Hasil Model Random Forest Regression**

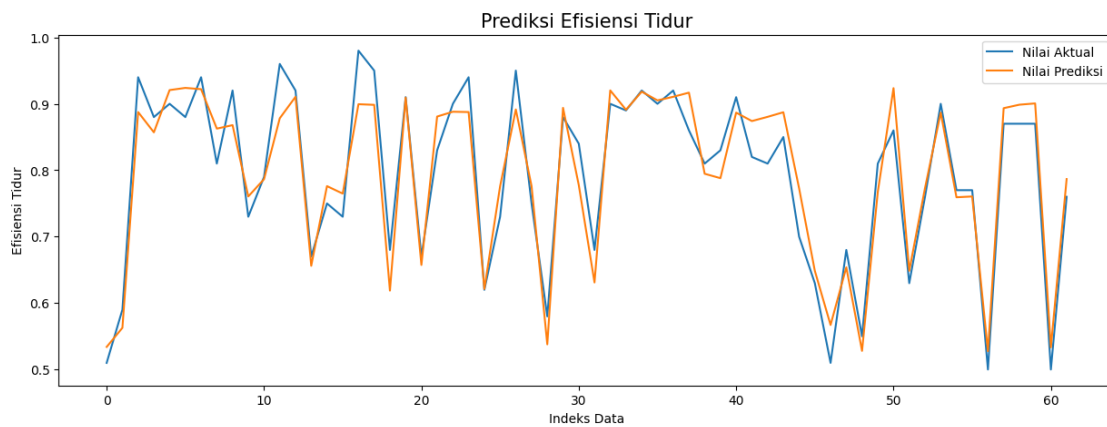
Gambar 6 menampilkan informasi bentuk pohon keputusan, Visualisasi ini menunjukkan bagaimana model memisahkan data berdasarkan nilai-nilai tertentu dari fitur input, seperti *light sleep percentage*, *awakenings*, *alcohol consumption*, dan variabel lainnya. Sementara itu, *leaf node* merepresentasikan nilai prediksi efisiensi tidur. Visualisasi ini memperlihatkan bagaimana model membentuk keputusan secara bertahap melalui kombinasi beberapa fitur penting dalam menghasilkan output numerik. Selanjutnya, model yang telah dilatih digunakan untuk membuat prediksi terhadap data uji ( $X\_test$ ). Prediksi yang dibuat dibandingkan dengan nilai sebenarnya ( $y\_test$ ) untuk mengevaluasi seberapa baik model mampu mempelajari pola dari data latih. Berikut ini adalah Tabel 2 yang memperlihatkan beberapa contoh hasil prediksi dibandingkan dengan nilai aktual.

**Tabel 2. Perbandingan Nilai Aktual dan Nilai Prediksi**

| Nilai Aktual | Nilai Prediksi |
|--------------|----------------|
| 0,51         | 0,539225       |
| 0,59         | 0,568960       |
| 0,94         | 0,889446       |
| 0,88         | 0,855786       |
| 0,90         | 0,919622       |
| ...          | ...            |
| 0,87         | 0,894270       |
| 0,87         | 0,903936       |
| 0,87         | 0,903309       |
| 0,50         | 0,535619       |
| 0,76         | 0,782954       |

Tabel 2 Perbandingan antara nilai aktual dan nilai prediksi menunjukkan bahwa model regresi yang digunakan memiliki performa yang cukup baik. Sebagian besar nilai prediksi berada sangat dekat dengan nilai aktual, dengan selisih yang relatif kecil. Misalnya, nilai aktual sebesar 0,90 diprediksi menjadi 0,919622, dan nilai aktual sebesar 0,87 diprediksi menjadi 0,894270. Hal ini

mencerminkan bahwa model mampu mengenali pola dari data secara efektif dan memberikan hasil prediksi yang akurat. Perbedaan yang kecil antara nilai aktual dan prediksi ini menjadi indikasi bahwa model memiliki kemampuan generalisasi yang baik terhadap data uji. Untuk memperkuat analisis, dibuat visualisasi perbandingan antara nilai aktual dan prediksi seperti ditampilkan pada Gambar 7.



**Gambar 7. Visualisasi Perbandingan Nilai Aktual dan Nilai Prediksi**

Melalui visualisasi perbandingan nilai aktual dan nilai prediksi yang telah dibuat, terdapat beberapa hal yang dapat diamati, yaitu pola sebaran prediksi. Pada grafik perbandingan gambar, terlihat bahwa model mengikuti tren dan pola nilai kualitas tidur aktual. Kemampuan model untuk mengidentifikasi pola umum data ditunjukkan oleh garis prediksi (oranye) yang secara umum mengikuti garis nilai aktual (biru). Terlihat juga pada bagian fluktuasi nilai efisiensi tidur menunjukkan fluktuasi nilai antara 0,5 dan 0,98, dengan mayoritas nilai berada di antara 0,7 dan 0,9. Model mampu memprediksi rentang fluktuasi ini dengan cukup baik. Namun tidak sedikit juga pada visualisasi ini, prediksi model tidak sepenuhnya sesuai dengan nilai aktual, terutama pada titik ekstrem (lokal maksimum dan minimum). Pada puncak-puncak tertentu, model cenderung memprediksi nilai yang lebih rendah dari aktual, sedangkan pada lembah-lembah tertentu, model cenderung memprediksi nilai yang lebih tinggi dari aktual.

**Evaluasi Model**

Setelah proses pemodelan selesai dan hasil prediksi diperoleh pada data pengujian, evaluasi performa model dilakukan untuk mengetahui seberapa baik model memprediksi nilai target. Evaluasi model ini dilakukan menggunakan empat metrik regresi, yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi ( $R^2$ ). Sebagai bentuk validasi tambahan, model *Random Forest Regression* dibandingkan dengan dua algoritma regresi lainnya, yaitu *Linear Regression* dan *Support Vector Regression* (SVR). Adapun hasil perbandingan ketiga model ditampilkan pada Tabel 3.

**Tabel 3. Perbandingan Kinerja Model Regresi dalam memprediksi Efisiensi Tidur**

| Model Regresi                    | MAE    | MSE    | RMSE   | $R^2$ |
|----------------------------------|--------|--------|--------|-------|
| <i>Random Forest Regression</i>  | 0,0327 | 0,0015 | 0,0390 | 0,91  |
| <i>Linear Regression</i>         | 0,0443 | 0,0029 | 0,0539 | 0,83  |
| <i>Support Vector Regression</i> | 0,0548 | 0,0042 | 0,0645 | 0,75  |

Berdasarkan nilai MAE sebesar 0.0327, MSE sebesar 0.0015, dan RMSE sebesar 0,0390, *Random Forest Regression* menunjukkan hasil terbaik berdasarkan tabel tersebut. Selain itu, model ini memiliki nilai  $R^2$  tertinggi sebesar 0,91, yang menunjukkan bahwa model dapat melakukan generalisasi terhadap data baru dengan baik tanpa mengalami overfitting yang signifikan. Hasil ini didukung oleh penelitian sebelumnya, seperti yang dilakukan oleh (Fadillah & Fauzan, 2024) dan

(Saragih et al., 2024) menunjukkan bahwa *Random Forest Regression* dapat menghasilkan nilai MAE dan MSE yang rendah dalam prediksi pada data yang kompleks, baik untuk prediksi kelayakan kredit maupun prevalensi stunting. (Özen, 2024) juga mencatat performa tinggi RFR dalam prediksi kasus COVID-19 dengan nilai  $R^2$  yang sangat tinggi. Sebaliknya, *Linear Regression* menunjukkan performa yang lebih rendah dengan MAE sebesar 0,0443 dan  $R^2$  sebesar 0,83. Meskipun model ini lebih sederhana dan mudah diinterpretasikan, hasil yang diperoleh sejalan dengan temuan (Patlisan, 2023) juga menunjukkan bahwa *Linear Regression*, meskipun lebih sederhana dan mudah diinterpretasikan, cenderung kurang efektif dalam menangani data yang kompleks. Sementara itu, *Support Vector Regression* (SVR) menghasilkan performa terendah dengan MAE sebesar 0,0548 dan  $R^2$  sebesar 0,75. Temuan ini bertolak belakang dengan studi (Penalun et al., 2023) yang menunjukkan keunggulan SVR dalam konteks prediksi laju penguapan. Namun, dalam studi ini, SVR kurang mampu mengenali pola non-linear dan kompleksitas dalam data efisiensi tidur. Secara keseluruhan, meskipun *Linear Regression* dan SVR lebih sederhana, *Random Forest Regression* terbukti lebih unggul dalam memodelkan data yang lebih kompleks dan menghasilkan prediksi yang lebih akurat.

### Visualisasi Hasil Prediksi

Setelah proses pelatihan dan evaluasi dilakukan, model *Random Forest Regression* menunjukkan performa yang baik dalam memproses data dan menghasilkan prediksi efisiensi tidur. Sistem kemudian digunakan untuk memprediksi efisiensi tidur berdasarkan data yang dimasukkan oleh pengguna melalui form input yang ditampilkan pada Gambar 8. Proses prediksi diawali dengan menganalisis faktor-faktor seperti usia, jenis kelamin, waktu tidur dan bangun, durasi tidur, jumlah terbangun saat malam, serta frekuensi olahraga mingguan. Temuan bahwa frekuensi olahraga berkorelasi positif dengan efisiensi tidur diperkuat oleh studi (He, 2023) dan (Li et al., 2021), yang menunjukkan bahwa aktivitas fisik rutin dapat meningkatkan kualitas tidur secara signifikan. Faktor lain yang turut dianalisis meliputi konsumsi kafein dan alkohol, status merokok, serta proporsi tidur REM, tidur ringan, dan tidur nyenyak.

The image shows two screenshots of a web application for sleep quality prediction. The left screenshot, titled "Prediksi Kualitas Tidur", is a form with the following fields: "Usia" (30), "Jenis Kelamin" (Laki-laki), "Jam Tidur (contoh: 22:30)" (01:00), "Jam Bangun (contoh: 06:00)" (06:00), "Durasi Tidur (contoh: 7 jam 30 menit - Input: 07:30)" (03:00), "Jumlah Terbangun Saat Tidur" (4), "Frekuensi Olahraga (per minggu)" (2), "Berapa kali Anda mengonsumsi kafein dalam seminggu?" (3), "Berapa kali Anda mengonsumsi alkohol dalam seminggu?" (1), "Merokok" (Ya), "Berapa jam tidur REM Anda?" (0), "Berapa jam tidur Ringan Anda?" (2), "Berapa jam tidur Nyenyak Anda?" (1), and "Merasa Kurang Istirahat?" (Ya). A "Prediksi" button is at the bottom. The right screenshot, titled "Hasil Prediksi Kualitas Tidur", shows "Efisiensi Tidur: 0.528" and "Kategori: Kurang". Below this, it says "Efisiensi tidur buruk. Disarankan untuk memperbaiki pola tidur dan mengurangi gangguan saat tidur." and a "Kembali ke Formulir" button.

**Gambar 8. Tampilan Sistem Prediksi Efisiensi Tidur dengan Random Forest Regression**

Setelah seluruh nilai inputan dimasukan, maka model sistem akan melakukan analisis seluruh data untuk mendapatkan hasil prediksi nilai efisiensi tidur menurut data melalui model *Random Forest Regression* yang telah dilatih sebelumnya. Model regresi ini menghasilkan output berupa nilai kontinu (desimal) rentang 0-1, bukan dalam bentuk kategori. Oleh karena itu, klasifikasi kualitas tidur seperti “Sangat Baik”, “Baik”, atau “Kurang” tidak dihasilkan langsung oleh model, melainkan ditentukan secara *post-hoc*, yaitu melalui proses interpretasi tambahan yang dilakukan setelah nilai prediksi diperoleh, dengan membandingkannya terhadap ambang batas klasifikasi yang telah

ditentukan. Sistem menghasilkan prediksi nilai efisiensi tidur sebesar 0,528, yang ekuivalen dengan 52,8% dalam bentuk persentase. Berdasarkan ambang klasifikasi efisiensi tidur dari (Reed & Sacco, 2016), yaitu *Sangat Baik* ( $\geq 95\%$ ), *Baik* (85–94%), dan *Kurang* ( $< 85\%$ ), hasil ini termasuk dalam kategori Kurang. Sistem kemudian memberikan umpan balik berupa saran perbaikan rutinitas tidur agar efisiensi tidur dapat ditingkatkan.

## SIMPULAN

Penelitian ini berhasil membangun model *Random Forest Regression* untuk memprediksi efisiensi tidur berdasarkan kombinasi faktor demografis dan kebiasaan harian. Model yang dibangun mampu mempelajari pola data dengan sangat baik melalui tahapan *preprocessing* yang menyeluruh dan pemilihan fitur yang tepat. Analisis korelasi menunjukkan bahwa persentase tidur nyenyak dan frekuensi olahraga memiliki hubungan positif terhadap efisiensi tidur, sementara jumlah terbangun menunjukkan korelasi negatif. Sebagai bentuk validasi tambahan, performa model ini dibandingkan dengan dua algoritma regresi lainnya, yaitu *Linear Regression* dan *Support Vector Regression* (SVR). Hasil perbandingan menunjukkan bahwa *Random Forest Regression* memberikan performa terbaik dibandingkan dua model lainnya, terutama dalam menangani data kompleks dan mengenali pola non-linear. Model dilatih dengan parameter *n\_estimators* sebanyak 100 dan *max\_depth* sebesar 5, serta dievaluasi menggunakan metrik regresi. Hasil evaluasi menunjukkan performa yang sangat baik, dengan nilai MAE sebesar 0,0327, MSE sebesar 0,0015, RMSE sebesar 0,0390, dan koefisien determinasi ( $R^2$ ) sebesar 0,91, yang menandakan kemampuan model dalam mengenali pola data dan melakukan generalisasi terhadap data uji.

Model kemudian diterapkan dalam sistem prediksi interaktif yang menerima input pengguna seperti usia, waktu tidur, durasi tidur, jumlah terbangun, dan kebiasaan harian lainnya. Salah satu hasil prediksi menunjukkan nilai efisiensi tidur sebesar 0,528. Nilai ini merupakan output model regresi dalam bentuk kontinu (0–1). Kategorisasi kualitas tidur seperti “Sangat Baik”, “Baik”, dan “Kurang” tidak dihasilkan langsung oleh model, melainkan dilakukan secara *post-hoc* dengan membandingkan hasil prediksi terhadap ambang batas yang telah ditentukan. Berdasarkan nilai ambang tersebut, prediksi sebesar 0,528 termasuk dalam kategori kualitas tidur “Kurang”. Oleh karena itu, penelitian ini menunjukkan bahwa kebiasaan harian dan komponen demografi memiliki pengaruh yang dapat diprediksi terhadap efisiensi tidur, dan dengan pendekatan *machine learning* yang dipakai di sini yaitu *Random Forest Regression* dapat menjadi alat yang efektif untuk menganalisis hubungan kompleks yang ada. Hasil ini mendukung penggunaan model prediktif dalam konteks pencegahan untuk meningkatkan kualitas tidur pengguna yang didasarkan pada data. Untuk pengembangan ke depan, eksplorasi algoritma lain dan tambahkan variabel seperti tingkat stres dan faktor lingkungan untuk analisis lanjutan.

## REFERENSI

- Altini, M., & Kinnunen, H. (2021). The promise of sleep: A multi-sensor approach for accurate sleep stage detection using the oura ring. *Sensors*, 21(13), 1–21. <https://doi.org/10.3390/s21134302>
- Ardesfira, G., Zedha, H. F., Fazana, I., Rahmadhiyanti, J., Rahima, S., & Anwar, S. (2022). Peramalan nilai tukar Rupiah terhadap Dollar Amerika dengan menggunakan metode Autoregressive Integrated Moving Average (ARIMA). *Jambura Journal of Probability and Statistics*, 3(2), 71–84. <https://doi.org/10.34312/jjps.v3i2.15469>
- Ardiansyah, D. (2023). Perbandingan model prediksi radiasi matahari berbasis mesin pembelajaran pada Stasiun Meteorologi Fatmawati Soekarno Bengkulu. *Megasains*, 14(1), 26–32. <https://doi.org/10.46824/megasains.v14i1.129>
- Cici Haryati, M., Kusumaningsih, I., & Supardi, S. (2022). Faktor-faktor yang berhubungan dengan kualitas tidur lansia di RW 03 Kelurahan Kebon Manggis Jakarta. *Jurnal Keperawatan Malang*, 7(2), 76–93. <https://jurnal.stikespantiwaluya.ac.id/index.php/JPW>
- Desjardins, S., Lapierre, S., Hudon, C., & Desgagné, A. (2019). Factors involved in sleep efficiency:

- A population-based study of community-dwelling elderly persons. *Sleep*, 42(5), 1–10. <https://doi.org/10.1093/sleep/zsz038>
- Driller, M. W., Dunican, I. C., Omond, S. E. T., Boukhris, O., Stevenson, S., Lambing, K., & Bender, A. M. (2023). Pyjamas, polysomnography and professional athletes: The role of sleep tracking technology in sport. *Sports*, 11(1), 1–17. <https://doi.org/10.3390/sports11010014>
- Fadillah, A. R., & Fauzan, M. N. (2024). Analisis perbandingan Linear Regression dan Random Forest Regression untuk prediksi batas kredit: Pendekatan optimasi hyperparameter. *Jurnal Informatika Polinema*, 10(4), 543–550. <https://doi.org/10.33795/jip.v10i4.5700>
- Faoziyah, N., & Suharjana, S. (2020). Kualitas dan efisiensi tidur yang buruk pada mahasiswa olahraga tahun pertama srata-1 terhadap pembelajaran gerak. *Jurnal SPORTIF : Jurnal Penelitian Pembelajaran*, 6(2), 514–525. [https://doi.org/10.29407/js\\_unpgri.v6i2.14545](https://doi.org/10.29407/js_unpgri.v6i2.14545)
- Farhanuddin, Sihombing, S. E. K., & Yahfizham. (2024). Komparasi multiple Linear Regression dan Random Forest Regression dalam memprediksi anggaran biaya manajemen proyek sistem informasi. *Journal of Computers and Digital Business*, 3(2), 86–97. <https://doi.org/10.56427/jcbd.v3i2.408>
- Hafizh, V., Putra, C., Al-husaini, M., Wahyu, A. P., & Raharja, A. R. (2025). Design of an intelligent monitoring system based on Internet of Things (IoT) with Random Forest Regression algorithm for height detection in cherry tomato plants. *Indonesian Journal of Machine Learning and Computer Science*, 5(January), 10–25. <https://doi.org/10.57152/malcom.v5i1.1612>
- Halson, S. L., Johnston, R. D., Piromalli, L., Lalor, B. J., Cormack, S., Roach, G. D., & Sargent, C. (2022). Sleep regularity and predictors of sleep efficiency and sleep duration in elite team sport athletes. *Sports Medicine - Open*, 8(1). <https://doi.org/10.1186/s40798-022-00470-7>
- He, Y. (2023). Effect of indicators of exercise on sleep efficiency. *Highlights in Science, Engineering and Technology*, 54(2), 342–350. <https://doi.org/10.1016/j.csm.2004.12.003>
- Li, L., Yu, Q., Zhao, W., Herold, F., Cheval, B., Kong, Z., Li, J., Mueller, N., Kramer, A. F., Cui, J., Pan, H., Zhan, Z., Hui, M., & Zou, L. (2021). Physical activity and inhibitory control: The mediating role of sleep quality and sleep efficiency. *Brain Sciences*, 11(5), 1–10. <https://doi.org/10.3390/brainsci11050664>
- Mazzotti, D. R., Guindalini, C., Moraes, W. A. dos S., Andersen, M. L., Cendoroglo, M. S., Ramos, L. R., & Tufik, S. (2014). Human longevity is associated with regular sleep patterns, maintenance of slow wave sleep, and favorable lipid profile. *Frontiers in Aging Neuroscience*, 6(JUN), 1–9. <https://doi.org/10.3389/fnagi.2014.00134>
- Meiranny, A., & Chabibah, A. M. (2022). Pengaruh konsumsi minuman berkafein terhadap pola dan kualitas tidur mahasiswa: A literatur review. *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, 5(2), 117–122. <https://doi.org/10.56338/mppki.v5i2.1910>
- Nugraha, Y. E. N., Ariawan, I., & Arifin, W. A. (2023). Weather forecast from time series data using LSTM algorithm. *Jurnal Teknologi Informasi Dan Komunikasi*, 14(1), 144–152. <https://doi.org/10.51903/jtikp.v14i1.531>
- Nur, A., Pudjianto, M., & Hidayat, E. Y. (2024). Perbandingan prediksi depresi mahasiswa dengan Linear Regression, Random Forest, dan Gradient Boosting. *SINTECH Journal*, 7(3), 180–189. <https://doi.org/10.31598/sintechjournal.v7i3.1729>
- Özen, F. (2024). Random forest regression for prediction of Covid-19 daily cases and deaths in Turkey. *Heliyon*, 10(4), 1–19. <https://doi.org/10.1016/j.heliyon.2024.e25746>
- Patlisan, P. (2023). Optimasi akurasi model Decision Tree menggunakan Random Forest Regression untuk prediksi kuantitas pembelian barang pada perusahaan manufaktur. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 14(2), 217–228. <https://doi.org/10.24176/simet.v14i2.9494>
- Penalun, F. E., Hermawan, A., & Avianto, D. (2023). Perbandingan Random Forest Regression dan Support Vector Regression pada prediksi laju penguapan. *Jurnal Fasilkom*, 13(02), 104–111. <https://doi.org/10.37859/jf.v13i02.4976>

- Prasani, M., Mutmainnah, M., & Mawarti, I. (2023). Gambaran kualitas tidur pada perempuan lansia di Puskesmas Simpang IV Sipin Kota Jambi. *Jurnal Ners*, 7(2), 1567–1575. <https://doi.org/10.31004/jn.v7i2.16687>
- Prastiyo, I. W., & Febriandirza, A. (2024). Analisis perbandingan prediksi tingkat kemiskinan menggunakan metode XGBoost dan Random Forest Regression. *Jurnal Media Informatika Budidarma*, 8(3), 1694. <https://doi.org/10.30865/mib.v8i3.7892>
- Reed, D. L., & Sacco, W. P. (2016). Measuring sleep efficiency: what should the denominator be? *Journal of Clinical Sleep Medicine*, 12(2), 263–266. <https://doi.org/10.5664/jcsm.5498>
- Ricky Verdiyanto, Dwi Hartanti, E. P. (2025). Pengembangan aplikasi point of sales untuk prediksi penjualan harian usaha minuman menggunakan algoritma Random Forest Regression. *Infotek : Jurnal Informatika Dan Teknologi*, 8(1), 128–139. <https://doi.org/10.29408/jit.v8i1.28386>
- Saragih, V. R., Arnita, A., Indra, Z., Taufik, I., & Sinaga, M. S. (2024). Comparison of supervised machine learning methods in predicting the prevalence of stunting in north sumatra province. *Journal of Soft Computing Exploration*, 5(4), 370–379. <https://doi.org/10.52465/josce.v5i4.498>
- Sehanguanaung, G. A., Mandey, S. L., & Roring, F. (2023). Analisis pengaruh harga, promosi dan kualitas pelayanan terhadap kepuasan konsumen pengguna aplikasi Lazada di Kota Manado. *Jurnal EMBA*, 11(3), 1–11. <https://iprice.co.id/insights/mapofecommerce/>
- Tjahjadi, M. E., Agustina, F. D., & Agnesta, R. (2022). Uji akurasi koordinat dari konfigurasi jaringan pemotretan menggunakan kamera non metrik: Studi kasus di Pandansari, Ngantang, Kabupaten Malang. *PROSIDING SEMSINA 2022*, 13(1), 81–89. <https://doi.org/10.36040/semsina.v3i1>
- Wulandari, S., & Pranata, R. (2024). Deskripsi kualitas tidur dan pengaruhnya terhadap konsentrasi belajar mahasiswa. *Jurnal Pendidikan Kesehatan Rekreasi*, 10(1), 101–108. <https://doi.org/10.59672/jpkr.v10i1.3414>