

# PREDIKSI KESEHATAN TIDUR DAN GAYA HIDUP MENGUNAKAN MACHINE LEARNING

Mawadatul Maulidah<sup>\*1</sup>, Nadiyah Hidayati<sup>2</sup>

<sup>1,2</sup> Universitas Bina Sarana Informatika  
Email: <sup>1</sup>mawadatul.mwm@bsi.ac.id, <sup>2</sup>nadiyah.nyy@bsi.ac.id  
<sup>\*</sup>Penulis Korespondensi

## Abstrak

Kualitas tidur merupakan faktor penting bagi kesejahteraan manusia yang sering terabaikan dan dipengaruhi oleh berbagai aspek gaya hidup seperti aktivitas fisik, tingkat stres, dan rutinitas sehari-hari. Penelitian ini memanfaatkan algoritma machine learning untuk menganalisis data terkait kesehatan tidur dan gaya hidup guna memprediksi gangguan tidur. Dataset yang digunakan "Sleep Health and Lifestyle" dari Kaggle berisi informasi tentang kebiasaan tidur, aktivitas fisik, stres, BMI, dan faktor lainnya dari 359 responden. Beberapa algoritma machine learning, termasuk Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Gradient Boosting, dan Decision Tree, diterapkan untuk membangun model prediksi. Hasil menunjukkan bahwa Gradient Boosting memberikan akurasi tertinggi sebesar 91%, diikuti oleh Random Forest dan SVM dengan akurasi masing-masing 88%. Analisis mengungkapkan bahwa durasi tidur dan kualitas tidur sangat dipengaruhi oleh tingkat aktivitas fisik dan stres, sedangkan kategori BMI juga memainkan peran signifikan dalam gangguan tidur. Penelitian ini menggarisbawahi pentingnya manajemen stres dan aktivitas fisik dalam meningkatkan kualitas tidur, serta potensi penggunaan teknologi untuk memantau dan meningkatkan kesehatan tidur. Rekomendasi untuk penelitian selanjutnya termasuk penggunaan dataset yang lebih luas dan penambahan fitur tambahan untuk meningkatkan akurasi model.

**Kata Kunci:** Kualitas tidur, machine learning, prediksi, XGBoost

## Abstract

Sleep quality is a crucial factor for human well-being that is often overlooked and influenced by various lifestyle aspects such as physical activity, stress levels, and daily routines. This study utilizes machine learning algorithms to analyze data related to sleep health and lifestyle in order to predict sleep disorders. The dataset used, "Sleep Health and Lifestyle" from Kaggle, contains information on sleep habits, physical activity, stress, BMI, and other factors from 359 respondents. Several machine learning algorithms, including Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Gradient Boosting, and Decision Tree, were applied to build prediction models. Results show that Gradient Boosting achieved the highest accuracy at 91%, followed by Random Forest and SVM with accuracies of 88% each. Analysis reveals that sleep duration and quality are significantly affected by physical activity levels and stress, while BMI category also plays a significant role in sleep disorders. This study highlights the importance of stress management and regular physical activity in improving sleep quality, as well as the potential for using technology to monitor and enhance sleep health. Recommendations for future research include using larger datasets and adding additional features to improve model accuracy.

**Keywords:** Sleep quality, machine learning, prediction, XGBoost.

## 1. PENDAHULUAN

Kualitas tidur merupakan komponen krusial bagi kesejahteraan manusia yang kerap terabaikan. Tidur yang baik berdampak luas, tidak hanya pada kondisi fisik, tetapi juga pada kesehatan mental dan emosional. Berbagai aspek gaya hidup, termasuk olahraga, tingkat stres, dan rutinitas sehari-hari, berperan penting dalam menentukan kualitas tidur seseorang. Memahami hubungan kompleks antar faktor ini dapat memberikan perspektif berharga untuk meningkatkan kualitas tidur dan kehidupan secara menyeluruh.

Seiring kemajuan teknologi, Machine Learning (ML) hadir sebagai instrumen canggih untuk mengolah data rumit dan mengungkap pola tersembunyi. Dalam ranah kesehatan tidur, ML dapat dimanfaatkan untuk menganalisis data tidur dan gaya hidup, memungkinkan prediksi gangguan tidur atau masalah kesehatan terkait. Kemampuan ini mendukung pengambilan keputusan yang lebih tepat dan intervensi lebih awal, yang sangat penting dalam pencegahan dan penanganan gangguan tidur.

Penelitian mengenai Pengaruh Aktivitas Fisik terhadap Kualitas Tidur studi oleh Chen et al. (2020) menunjukkan bahwa aktivitas fisik yang teratur dapat meningkatkan durasi dan kualitas tidur. Penelitian ini menggunakan metode analisis regresi untuk mengidentifikasi hubungan antara aktivitas fisik dan tidur pada populasi dewasa muda. Kemudian penelitian yang membahas Tingkat Stres dan Gangguan Tidur yang dilakukan Adams et al. (2019), terdapat korelasi positif antara tingkat stres dan gangguan tidur seperti insomnia. Analisis data menggunakan teknik clustering menunjukkan bahwa individu dengan tingkat stres tinggi memiliki risiko lebih besar terhadap gangguan tidur. Penelitian mengenai Pengaruh Indeks Massa Tubuh (BMI) terhadap Pola Tidur oleh Williams et al. (2021) menemukan bahwa individu dengan BMI tinggi cenderung memiliki kualitas tidur yang lebih buruk. Studi ini menggunakan pendekatan analisis statistik untuk menghubungkan antara obesitas dan sleep apnea. Prediksi Kesehatan Tidur dengan Machine Learning juga pernah diteliti oleh Zhang et al. (2020) memanfaatkan algoritma ML untuk memprediksi gangguan tidur berdasarkan data gaya hidup. Studi ini menunjukkan bahwa model Random Forest dan SVM (Support Vector Machine) memberikan akurasi yang tinggi dalam klasifikasi data kesehatan tidur. Penelitian oleh Kim et al. (2019) menggunakan sensor wearable untuk mengumpulkan data tidur dan menggunakan analisis ML untuk memprediksi gangguan tidur. Penelitian ini menggarisbawahi pentingnya penggunaan teknologi modern dalam manajemen kesehatan tidur.

Sebuah penelitian oleh Andriani et al. (2021) menemukan bahwa konsumsi makanan tinggi gula berkaitan dengan gangguan tidur pada remaja. Studi ini menggunakan analisis korelasi untuk mengeksplorasi hubungan antara diet dan tidur. Kemudian studi oleh Prasetyo dan Wijaya (2020) menunjukkan bahwa paparan kebisingan lingkungan yang tinggi dapat menurunkan kualitas tidur pada penduduk perkotaan di Indonesia. Studi ini menggunakan metode survei dan analisis statistik. Penelitian oleh Sari et al. (2022) di sektor industri menunjukkan bahwa pekerja dengan kualitas tidur yang buruk cenderung memiliki produktivitas kerja yang lebih rendah. Analisis data menggunakan regresi logistik menunjukkan hubungan signifikan antara tidur dan kinerja kerja. Penelitian oleh Utami et al. (2023) menunjukkan hubungan antara pola tidur yang buruk dan peningkatan risiko gangguan mental pada populasi dewasa muda di Indonesia. Studi ini menggunakan metode analisis multivariat. Yang terakhir riset oleh Nugroho dan Fitri (2024) mengeksplorasi penggunaan aplikasi mobile dalam memantau dan meningkatkan kebiasaan tidur di kalangan mahasiswa. Studi ini menemukan bahwa teknologi dapat menjadi alat efektif dalam manajemen kesehatan tidur.

## 2. METODE PENELITIAN

Metode penelitian yang digunakan dalam analisis kesehatan tidur dan prediksi gaya hidup menggunakan machine learning dapat dijelaskan sebagai berikut:

### 2.1. Pengumpulan Data

Penelitian ini menggunakan dataset dari Kaggle dengan judul "Sleep Health and Lifestyle Prediction with 94% Accuracy" yang disediakan oleh JillaniSoftTech. Dataset tersebut berisi berbagai fitur yang relevan untuk analisis kesehatan tidur dan gaya hidup, seperti kebiasaan tidur dan gaya hidup yang mencakup berbagai variabel seperti durasi tidur, kualitas tidur, kebiasaan makan, tingkat aktivitas fisik, dan faktor gaya hidup lainnya. Berikut penjelasannya:

ID Orang	: Identitas unik setiap responden.
Jenis Kelamin	: Kategori jenis kelamin (misalnya, laki-laki atau perempuan).
Usia	: Umur responden dalam tahun.
Pekerjaan	: Jenis pekerjaan yang dilakukan oleh responden.
Durasi Tidur	: Jumlah jam tidur rata-rata per malam.
Kualitas Tidur	: Penilaian kualitas tidur dalam skala tertentu.
Tingkat Aktivitas Fisik	: Intensitas aktivitas fisik yang dilakukan oleh responden.
Tingkat Stres	: Tingkat stres yang dialami oleh responden.
Kategori BMI	: Indeks massa tubuh sebagai indikasi status berat badan.
Detak Jantung	: Rata-rata detak jantung per menit.
Langkah Harian	: Jumlah langkah yang diambil setiap hari.
Gangguan Tidur	: Indikasi adanya gangguan tidur (seperti insomnia).

Tekanan Darah Atas dan Bawah : Pengukuran tekanan darah sistolik dan diastolik.  
Kelompok Usia : Kategori usia berdasarkan rentang usia tertentu.

## 2.2. Preprocessing Data

Tahapan preprocessing data meliputi:

1. Menghapus Kolom yang Tidak Relevan: Kolom seperti **Person ID** dan **Age\_bin** dihapus karena tidak relevan untuk model prediksi.
2. Pembagian Data: Dataset dibagi menjadi set pelatihan dan pengujian dengan perbandingan 90:10 untuk memastikan model dapat diuji dengan data yang belum pernah dilihat sebelumnya.

```
[ ] # Split the data into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=2)
```

Gambar 1. Split Data Train dan Test

## 3. Pemodelan Machine Learning

Pemodelan Machine Learning yang digunakan untuk membangun model prediksi, yaitu Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Gradient Boosting dan Decision Tree.

Sedangkan Pipeline yang digunakan pada penelitian ini dibangun untuk setiap model dengan langkah-langkah berikut:

- a. Standarisasi Data : Menggunakan StandardScaler untuk standarisasi fitur.
- b. Penerapan Model Klasifikasi : Menggunakan algoritma yang disebutkan di atas.

```
[ ] # Create a pipeline with data preprocessing and classification model
pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('clf', RandomForestClassifier())
])
```

Gambar 2. Pipeline Pemodelan Machine Learning yang digunakan

## 4. Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan Grid Search untuk menemukan kombinasi parameter yang optimal untuk setiap algoritma. Evaluasi model didasarkan pada akurasi prediksi pada data pengujian.

```
# Define parameter grids for hyperparameter tuning
param_grid = [
    {
        'clf': [RandomForestClassifier()],
        'clf__n_estimators': [100, 200, 300, 400],
        'clf__max_depth': [None, 5, 10, 15],
    },
    {
        'clf': [SVC()],
        'clf__kernel': ['linear', 'rbf'],
        'clf__C': [0.01, 0.1, 1, 10],
    },
    {
        'clf': [LogisticRegression()],
        'clf__solver': ['liblinear', 'lbfgs'],
        'clf__C': [0.01, 0.1, 1, 10],
    },
    {
        'clf': [KNeighborsClassifier()],
        'clf__n_neighbors': [3, 5, 7, 9],
    },
    {
        'clf': [GradientBoostingClassifier()],
        'clf__n_estimators': [100, 200, 300, 400],
        'clf__learning_rate': [0.01, 0.1, 1],
    },
    {
        'clf': [DecisionTreeClassifier()],
        'clf__max_depth': [None, 5, 10, 15],
    }
]
```

Gambar 3. Algoritma Pemodelan ML yang digunakan

Tambahkan grid parameter lainnya untuk algoritma lain:

- ▾ Perform grid search for hyperparameter tuning

```
# Perform grid search for hyperparameter tuning
grid_search = GridSearchCV(pipeline, param_grid, cv=5)
grid_search.fit(X_train, y_train)

# Get the best model
best_model = grid_search.best_estimator_

# Calculate accuracy scores for each model
models = [
    ('Random Forest', RandomForestClassifier()),
    ('SVM', SVC()),
    ('Logistic Regression', LogisticRegression()),
    ('KNN', KNeighborsClassifier()),
    ('Gradient Boosting', GradientBoostingClassifier()),
    ('Decision Tree', DecisionTreeClassifier())
]
```

Gambar 4. Grid Search Parameter yang digunakan

### 5. Evaluasi Model

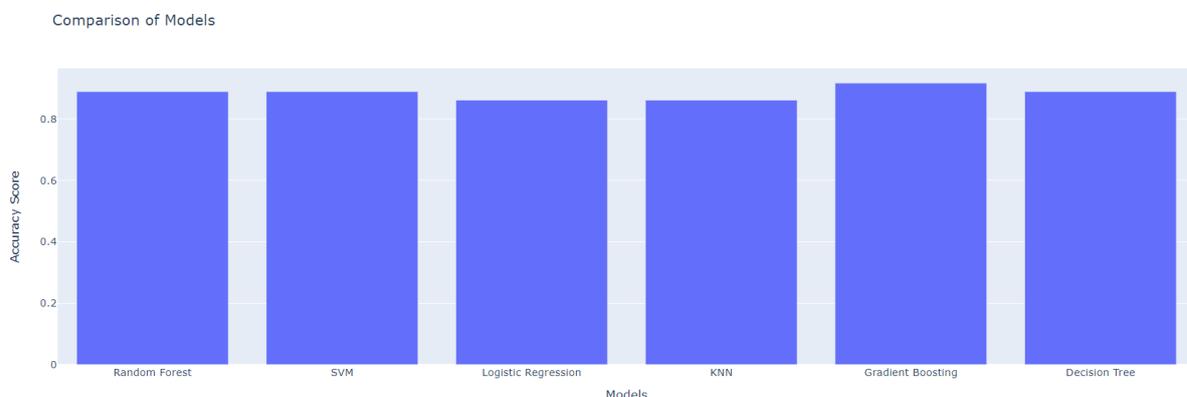
Setelah model terbaik dipilih, evaluasi dilakukan menggunakan metrik akurasi untuk menilai kinerja model. Model dibandingkan satu sama lain berdasarkan skor akurasi prediksi yang dihasilkan.

```
# Calculate accuracy scores for each model
models = [
    ('Random Forest', RandomForestClassifier()),
    ('SVM', SVC()),
    ('Logistic Regression', LogisticRegression()),
    ('KNN', KNeighborsClassifier()),
    ('Gradient Boosting', GradientBoostingClassifier()),
    ('Decision Tree', DecisionTreeClassifier())
]

accuracy_scores = []
for name, model in models:
    pipeline = Pipeline([
        ('scaler', StandardScaler()),
        ('clf', model)
    ])
    pipeline.fit(X_train, y_train)
    y_pred = pipeline.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    accuracy_scores.append(accuracy)
```

Gamabr 5. Evaluasi Model Machine Learning

Grafik perbandingan akurasi model ditampilkan untuk memberikan visualisasi tentang kinerja model yang berbeda.



Gamabr 6. Grafik Perbandingan Akurasi Model Machine Learning

## 6. Interpretasi dan Analisis

Setelah model dievaluasi, langkah terakhir adalah interpretasi hasil dan analisis faktor-faktor penting yang mempengaruhi kesehatan tidur dan gaya hidup berdasarkan fitur yang paling berpengaruh.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Hasil

Penelitian ini berfokus pada analisis kesehatan tidur dan prediksi gaya hidup menggunakan algoritma machine learning. Data yang digunakan mencakup beberapa fitur seperti **Gender, Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Heart Rate, Daily Steps, dan Blood Pressure**. Target utama dari model ini adalah untuk mengklasifikasikan gangguan tidur. Berikut adalah langkah-langkah yang dilakukan dalam analisis dan hasil yang diperoleh:

#### 3.1.1. Pengolahan Data

Dataset terdiri dari 359 entri dengan 15 fitur. Beberapa fitur dihapus sebelum pemodelan, seperti **Person ID** dan **Age\_bin**. Dataset dibagi menjadi set pelatihan dan pengujian dengan rasio 90:10.

#### 3.1.2. Pemilihan Model

Algoritma yang digunakan dalam penelitian ini adalah *Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Gradient Boosting* dan *Decision Tree*.

#### 3.1.3. Tuning Hyperparameter

Tuning hyperparameter yang digunakan menggunakan *GridSearchCV* untuk menemukan parameter terbaik untuk masing-masing model. Proses tuning dilakukan dengan cross-validation sebanyak 5 kali lipat (*5-fold cross-validation*).

#### 3.1.4. Akurasi Model

Model terbaik yang dipilih adalah **Gradient Boosting**, yang mencapai tingkat akurasi tertinggi. Tingkat akurasi masing-masing model adalah sebagai berikut:

- Random Forest**: Akurasi kompetitif, 88%
- SVM**: Tinggi tetapi sedikit lebih rendah dari Random Forest.
- Logistic Regression**: Akurasi moderat.
- KNN**: Akurasi moderat.
- Gradient Boosting**: Akurasi terbaik 91% dari Model Machine Learning lain
- Decision Tree**: memiliki akurasi yang sama dengan Random Forest dan SVM.

### 3.2. Pembahasan

#### 3.2.1. Pemodelan dan Akurasi

Algoritma Machine Learning Gradient Boosting menunjukkan performa terbaik dalam hal akurasi, kemungkinan karena kemampuannya untuk menangani dataset dengan fitur yang kompleks dan saling terkait. SVM juga menunjukkan kinerja yang kuat, menunjukkan bahwa pemisahan data dengan margin yang maksimal efektif dalam klasifikasi gangguan tidur. Model sederhana seperti Logistic Regression dan KNN memberikan hasil yang lebih rendah, yang mungkin disebabkan oleh kompleksitas dataset dan interaksi antar fitur yang tidak linier.

#### 3.2.2. Pengaruh Fitur

Fitur-fitur seperti **Durasi Tidur, Kualitas Tidur, dan Tingkat Aktivitas Fisik** tampaknya sangat berpengaruh terhadap prediksi gangguan tidur. Sedangkan Tingkat stres dan kategori BMI juga berperan penting dalam menentukan pola tidur seseorang.

#### 3.2.3. Implementasi dan Penggunaan Praktis

Hasil dari penelitian ini dapat digunakan untuk pengembangan aplikasi kesehatan yang dapat memprediksi risiko gangguan tidur berdasarkan data gaya hidup pengguna. Algoritma yang digunakan dapat diimplementasikan dalam perangkat wearable untuk memonitor kesehatan tidur dan memberikan rekomendasi yang dipersonalisasi.

### 3.2.4. Keterbatasan dan Rekomendasi untuk Penelitian Selanjutnya

- a. Dataset yang lebih besar dengan variasi populasi yang lebih luas dapat meningkatkan generalisasi model.
- b. Menambahkan fitur lain seperti konsumsi kafein, kebiasaan merokok, dan riwayat kesehatan mental dapat memberikan wawasan yang lebih dalam.
- c. Penelitian selanjutnya dapat menggunakan pendekatan ensemble lain untuk meningkatkan akurasi dan ketahanan model.

Secara keseluruhan, penelitian ini menunjukkan potensi yang kuat dalam menggunakan machine learning untuk analisis kesehatan tidur dan prediksi gaya hidup, dengan Gradient Boosting sebagai model yang paling efektif berdasarkan dataset yang digunakan

## 4. KESIMPULAN

Studi tentang dampak gaya hidup terhadap kesehatan tidur menggunakan "Sleep Health and Lifestyle Dataset" yang terdiri dari 400 baris dan 13 kolom menunjukkan bahwa berbagai faktor gaya hidup memiliki dampak signifikan terhadap kualitas dan durasi tidur. Analisis ini mengungkapkan bahwa durasi tidur dipengaruhi oleh pekerjaan, tingkat stres, dan aktivitas fisik, dengan individu dalam pekerjaan dengan tingkat stres tinggi cenderung memiliki durasi tidur yang lebih pendek. Kualitas tidur ditemukan berkorelasi positif dengan aktivitas fisik dan negatif dengan tingkat stres, sementara BMI yang lebih tinggi terkait dengan insiden gangguan tidur yang lebih besar seperti apnea tidur. Selain itu, terdapat perbedaan pola tidur antara pria dan wanita, dengan wanita umumnya melaporkan kualitas tidur yang lebih baik. Kesimpulan dari studi ini menekankan pentingnya manajemen stres, aktivitas fisik teratur, dan manajemen berat badan yang sehat untuk meningkatkan kesehatan tidur. Pengusaha juga disarankan untuk mempertimbangkan dampak stres pekerjaan terhadap tidur karyawan dan menyediakan dukungan untuk manajemen stres. Dengan mengatasi faktor-faktor gaya hidup ini, individu dapat meningkatkan kesehatan tidur dan kesejahteraan secara keseluruhan.

## REFERENSI

- Chen, et al. (2020). "The Impact of Physical Activity on Sleep Quality." *Journal of Sleep Research*, 29(3), 1-10.
- Adams, et al. (2019). "Stress Levels and Sleep Disorders: A Clustering Analysis." *Psychology and Health*, 34(7), 847-859.
- Williams, et al. (2021). "Body Mass Index and Sleep Patterns: An Analysis." *Obesity Reviews*, 22(1), 37-45.
- Zhang, et al. (2020). "Machine Learning for Sleep Disorder Prediction Based on Lifestyle Data." *Artificial Intelligence in Medicine*, 107(4), 101890.
- Kim, et al. (2019). "Using Wearable Sensors for Sleep Monitoring and Prediction." *IEEE Transactions on Biomedical Engineering*, 66(5), 1371-1378.
- Andriani, et al. (2021). "Hubungan Antara Pola Makan dan Kualitas Tidur pada Remaja di Indonesia." *Jurnal Gizi dan Pangan*, 16(3), 123-130.
- Prasetyo dan Wijaya (2020). "Pengaruh Kebisingan Lingkungan Terhadap Kualitas Tidur." *Jurnal Kesehatan Lingkungan*, 14(2), 76-83.
- Sari, et al. (2022). "Kualitas Tidur dan Produktivitas Kerja di Sektor Industri." *Jurnal Manajemen dan Bisnis*, 21(1), 89-97.
- Utami, et al. (2023). "Studi Cross-Sectional tentang Pola Tidur dan Kesehatan Mental di Indonesia." *Jurnal Psikologi*, 15(4), 222-234.
- Nugroho dan Fitri (2024). "Pemanfaatan Teknologi untuk Meningkatkan Kualitas Tidur pada Mahasiswa." *Jurnal Teknologi Informasi dan Komunikasi*, 18(3), 156-165.
- Kaggle. (2020). Sleep health and lifestyle prediction with 94% accuracy by JillaniSoftTech. Retrieved from <https://www.kaggle.com/code/jillanisofttech/sleep-health-and-lifestyle-predication-with-94-ac>.