

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Model Prediksi Gangguan Tidur berdasarkan Beberapa Faktor menggunakan
Machine Learning**

**Prediction of Sleep Disorders based on Several Factors using Machine
Learning**

Faradillah^{1*}, Muhammad Fadhiel Alie², Rian Rahmanda³

^{1,2} Program Studi Sistem Informasi, Universitas Indo Global Mandiri
Jalan Sudirman No. 629 Kota Palembang

³ Program Studi Sistem Komputer, Politeknik Negeri Sriwijaya
Jalan Srijaya Negara Bukit Lama Palembang

email: ¹faradillah.hakim@uigm.ac.id, ²fadhiel@uigm.ac.id, ³rianrahmanda@polsri.ac.id

ABSTRAK

Gangguan tidur merupakan masalah kesehatan yang signifikan dan dapat mempengaruhi kualitas hidup individu. Penelitian ini bertujuan untuk mengembangkan model prediksi gangguan tidur menggunakan teknik *machine learning* dengan mempertimbangkan beberapa faktor risiko. Dataset yang digunakan merupakan data sekunder yang diperoleh dari *Kaggle* dan dianalisis menggunakan beberapa algoritma machine learning, termasuk model *machine learning* seperti *logistic regression*, *decision tree* dan *gradient boosting*. Hasil penelitian menunjukkan bahwa model Gradient Boosting menghasilkan akurasi prediksi tertinggi, dengan nilai akurasi 99% berdasarkan AUC – ROC Score. Faktor-faktor seperti usia (*ages*), durasi tidur (*sleep duration*), kategori BMI dan pekerjaan ditemukan sebagai prediktor yang paling signifikan. Temuan ini menunjukkan bahwa penggunaan machine learning dapat menjadi alat yang efektif dalam mengidentifikasi individu yang berisiko mengalami gangguan tidur, sehingga memungkinkan intervensi dini dan pengelolaan kesehatan yang lebih baik. Penelitian ini memberikan kontribusi penting dalam pemahaman tentang hubungan antara faktor-faktor risiko dan gangguan tidur serta potensi aplikasi *machine learning* dalam bidang melalui pemilihan model prediksi dengan akurasi terbaik.

Kata Kunci: Gangguan tidur; *Machine learning*; Prediksi.

ABSTRACT

Sleep disorders are a significant health problem and can affect an individual's quality of life. This research aims to develop a prediction model for sleep disorders using machine learning techniques by considering several risk factors. The dataset used is secondary data obtained from *Kaggle* and analyzed using several machine learning algorithms, including machine learning models such as *logistic regression*, *decision trees* and *gradient boosting*. The research results show that the Gradient Boosting model produces the highest prediction accuracy, with

*Penulis Korespondensi: Faradillah
email: faradillah.hakim@uigm.ac.id

an accuracy value of 99% based on AUC – ROC Score. Factors such as age (ages), sleep duration, BMI category and occupation were found to be the most significant predictors. These findings suggest that the use of machine learning can be an effective tool in identifying individuals at risk of sleep disorders, allowing for early intervention and better health management. This research makes an important contribution to the understanding of the relationship between risk factors and sleep disorders as well as the potential application of machine learning in the field through the selection of prediction models with the best accuracy.

Keywords: *Sleep disorders; Machine learning; Prediction.*

1. PENDAHULUAN

Gangguan tidur adalah kondisi abnormal yang dapat mengakibatkan permasalahan dalam siklus tidur seseorang [1][2]. Gangguan tidur yang terjadi secara terus menerus dapat memiliki dampak serius pada kesehatan fisik, mental, dan kualitas hidup seseorang[3][4]. Kurang tidur dapat menyebabkan kelelahan terus-menerus dan penurunan energi sepanjang hari[1]. Hal tersebut juga dapat melemahkan sistem kekebalan tubuh, membuat seseorang lebih rentan terhadap infeksi dan penyakit. Gangguan tidur, terutama sleep apnea, dapat meningkatkan risiko tekanan darah tinggi, penyakit jantung, dan stroke[4][5]. Kurang tidur dapat mempengaruhi hormon yang mengatur nafsu makan, sehingga meningkatkan risiko obesitas. Selain itu Kurang tidur dapat mengganggu metabolisme glukosa dan meningkatkan risiko diabetes tipe 2. Gangguan tidur dapat memperburuk atau memicu kondisi kesehatan mental seperti depresi dan kecemasan. Kurang tidur dapat mempengaruhi fungsi otak, mengakibatkan kesulitan dalam konsentrasi, pengambilan keputusan, dan masalah memori. Orang yang kurang tidur cenderung mengalami perubahan suasana hati, iritabilitas, dan stres. Kurang tidur dapat mempengaruhi produktivitas dan kinerja di tempat kerja atau sekolah. Mengantuk di siang hari dapat meningkatkan risiko kecelakaan, baik di jalan raya maupun di tempat kerja. Gangguan tidur dapat mempengaruhi hubungan sosial dan menyebabkan isolasi sosial karena perubahan suasana hati dan kelelahan[4][5][6][7][8]. Hal ini dapat mengganggu kualitas dan menurunkan produktivitas hidup seseorang [9][10][11][12]. Kualitas tidur dapat dipengaruhi oleh beberapa faktor seperti durasi tidur, level aktivitas fisik, level stress, kategori indeks massa tubuh (BMI), tekanan darah, detak jantung dan jumlah langkah setiap hari [10][11][12][13]. Menurut Organisasi Kesehatan Dunia (WHO), sekitar 30% populasi dunia menderita beberapa bentuk gangguan tidur. Di Indonesia, prevalensi gangguan tidur juga cukup tinggi, dan menurut penelitian, sekitar 20-30% orang dewasa menderita gangguan tidur kronis seperti insomnia, sleep apnea, atau restless leg syndrome[14][15][16].

Beberapa penelitian telah dilakukan untuk memprediksi penyakit berdasarkan berbagai faktor. Gao melakukan tinjauan sistematis terhadap meta-analisis pada penelitiannya. Meta-analisis ini mengevaluasi hubungan antara durasi dan kualitas tidur serta berbagai hasil kesehatan. Hasil penelitian Gao menunjukkan bahwa durasi tidur yang lama dikaitkan dengan peningkatan risiko kematian[9]. Kualitas tidur yang buruk juga dikaitkan dengan peningkatan risiko diabetes dan diabetes gestasional[9]. Studi lain (Arora, Chakraborty, dan Bhatia 2020) menggunakan pembelajaran mendalam untuk memprediksi kualitas tidur berdasarkan data dari perangkat seluler. Mereka menemukan bahwa tiga indikator tidur dan CNN lebih akurat dibandingkan MLP dengan akurasi 97,30%[8]. Sebuah studi oleh Journal of Medical Internet

Research Group yang dilakukan oleh Seokmin Ha dkk menunjukkan bahwa membedakan antara COMISA dan OSA penting untuk memprediksi gangguan tidur secara akurat. Mereka menciptakan algoritma untuk memprediksi risiko gangguan tidur dan menunjukkan bagaimana risiko tersebut berubah seiring dengan perubahan berat badan atau usia[17].

Mengingat dampak luas dari gangguan tidur, penting untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap kondisi ini dan mengembangkan metode untuk mendeteksinya secara dini. Pada tulisan ini akan membuat model prediksi ada atau tidaknya gangguan tidur berdasarkan faktor-faktor tersebut menggunakan *machine learning* yang dapat memberikan wawasan dan membantu dalam deteksi dini dan penanganan gangguan tidur. Pada tulisan ini akan diimplementasikan beberapa model *machine learning* seperti *logistic regression*, *decision tree* dan *gradient boosting*. Dari ketiga model tersebut, akan dilihat mana yang memiliki kinerja yang paling baik dalam memprediksi gangguan tidur berdasarkan metrik akurasi, presisi, recall, F1-score dan AUC-ROC Score.

2. METODE

Penelitian ini dilalui beberapa tahapan yaitu tahap persiapan data, pemilihan features, pemilihan model, dan prediksi gangguan tidur menggunakan model yang terpilih[18][19]. Penelitian ini menggunakan pendekatan machine learning untuk memprediksi gangguan tidur berdasarkan beberapa faktor seperti durasi tidur, kualitas tidur, tingkat aktivitas fisik, tingkat stres, kategori BMI, tekanan darah, detak jantung, dan jumlah langkah harian. Dataset yang digunakan dalam penelitian ini diambil dari Kaggle yang merupakan platform terkemuka yang menyediakan berbagai dataset berkualitas untuk keperluan analisis data dan machine learning. Dataset yang bersumber dari Kaggle mudah diakses dan umumnya gratis untuk digunakan. Hal ini memudahkan peneliti untuk mendapatkan data yang diperlukan tanpa hambatan biaya atau akses. Kaggle menawarkan berbagai macam dataset dari berbagai domain, termasuk kesehatan, yang relevan dengan penelitian gangguan tidur.

Dataset di Kaggle telah divalidasi dan digunakan dalam berbagai kompetisi dan proyek penelitian, menunjukkan tingkat kualitas yang tinggi. Selain itu, dataset ini sering dilengkapi dengan dokumentasi yang baik, memudahkan pemahaman dan penggunaan data. Evaluasi kualitas data dilakukan melalui beberapa langkah berikut: Pemeriksaan Kelengkapan Data, Memastikan bahwa dataset tidak memiliki terlalu banyak nilai yang hilang (missing values), Mengevaluasi konsistensi dan validitas data dengan melakukan pemeriksaan logis terhadap nilai-nilai yang ada, Melakukan pembersihan data untuk mengatasi outliers, data duplikat, dan nilai yang tidak valid, serta melakukan normalisasi data jika diperlukan, Melakukan analisis deskriptif untuk memahami distribusi dan karakteristik data, yang dapat membantu dalam memahami konteks dan relevansi data terhadap penelitian.

Dalam penelitian ini, beberapa algoritma machine learning dipilih untuk membangun model prediksi, yaitu Logistic Regression, Decision Tree, dan Gradient Boosting. Logistic Regression merupakan algoritma yang sederhana dan mudah diinterpretasikan sehingga mudah memahami hubungan antara variabel input dan output. Baseline Model Logistic Regression sering digunakan sebagai model baseline untuk membandingkan performa model yang lebih kompleks. Sementara Decision Tree mampu menangani hubungan non-linear antara variabel. Decision Tree juga cukup mudah diinterpretasikan, memberikan wawasan tentang bagaimana keputusan diambil berdasarkan variabel input. Gradient Boosting sering disebut High Performance, Gradient Boosting sering memberikan performa yang sangat baik dalam berbagai kompetisi machine learning, karena mampu menangani kompleksitas data

dengan baik. Handling Overfitting, Teknik boosting membantu dalam mengurangi overfitting, meningkatkan generalisasi model. Proses pemilihan model dilakukan melalui langkah-langkah berikut:

- **Data Splitting:** Dataset dibagi menjadi training set dan testing set untuk melatih model dan menguji performa model.
- **Model Training:** Melatih beberapa model menggunakan training set. Model yang dilatih meliputi Logistic Regression, Decision Tree, dan Gradient Boosting.
- **Cross-validation:** Melakukan cross-validation untuk memastikan bahwa model tidak overfit terhadap data training.
- **Performance Evaluation:** Mengukur performa model menggunakan metrik seperti: accuracy, precision, recall, dan F1-score pada data testing. Membandingkan performa model berdasarkan metrik-metrik tersebut.
- **Model Selection:** Memilih model dengan performa terbaik berdasarkan hasil evaluasi. Model yang dipilih adalah yang memiliki keseimbangan terbaik antara akurasi dan kemampuan generalisasi. Dengan pendekatan ini, penelitian diharapkan dapat menghasilkan model prediksi gangguan tidur yang akurat dan dapat diandalkan, yang dapat digunakan untuk intervensi dan pengambilan keputusan yang lebih baik dalam menangani masalah gangguan tidur.

2.1 Tahapan Persiapan Data

Dataset yang digunakan pada tulisan ini didapatkan dari Kaggle[20]. Dataset yang digunakan dalam penelitian ini diambil dari Kaggle, dataset ini mencakup berbagai jenis data yang relevan untuk prediksi gangguan tidur, termasuk data demografis, data kesehatan, dan data aktivitas fisik. Tipe data yang terdapat dalam dataset ini meliputi: (a) Numerik yaitu Durasi tidur (jam), tekanan darah (mmHg), detak jantung (bpm), jumlah langkah harian; (b) Kategorikal yaitu Kualitas tidur (baik, cukup, buruk), tingkat stres (rendah, sedang, tinggi), kategori BMI (kurus, normal, overweight, obesitas). Dataset ini terdiri dari sekitar 5000 sampel. Setiap sampel mewakili data seorang individu dengan informasi yang lengkap tentang faktor-faktor yang dianggap relevan untuk penelitian gangguan tidur. Dataset ini sangat relevan dengan tujuan penelitian karena mencakup berbagai faktor yang telah diidentifikasi dalam literatur sebagai penentu penting kualitas tidur dan risiko gangguan tidur. Faktor-faktor ini termasuk durasi tidur, kualitas tidur, tingkat aktivitas fisik, tingkat stres, BMI, tekanan darah, detak jantung, dan jumlah langkah harian.

Dalam penelitian ini, penanganan data yang hilang dilakukan dengan cermat untuk memastikan integritas data dan keandalan model prediksi. Metode yang digunakan antara lain: Pemeriksaan Kelengkapan Data, Metode Imputasi, Untuk kolom numerik, nilai yang hilang diimputasi menggunakan metode mean. Metode ini dipilih karena sederhana dan efektif dalam mengurangi bias yang mungkin timbul akibat nilai yang hilang. Misalnya, jika ada nilai yang hilang pada kolom durasi tidur, nilai rata-rata durasi tidur dari sampel yang tersedia digunakan untuk mengisi nilai yang hilang tersebut. Kategorikal, untuk kolom kategorikal, nilai yang hilang diimputasi menggunakan metode mode (nilai yang paling sering muncul). Ini membantu dalam mempertahankan distribusi kategori yang ada dalam data. Misalnya, jika ada nilai yang hilang pada kolom kualitas tidur, kategori yang paling sering muncul (misalnya, 'cukup') digunakan untuk mengisi nilai yang hilang tersebut. Selanjutnya dalam penelitian ini, metode encoding kategorikal digunakan untuk mengubah data kategorikal menjadi format yang dapat diproses oleh algoritma machine learning.

Dua metode yang digunakan adalah *one-hot encoding* kolom kategorikal yang tidak memiliki urutan inherent, seperti kualitas tidur dan tingkat stres. One-hot encoding mengubah setiap kategori menjadi kolom biner terpisah. One-hot encoding mencegah algoritma dari menafsirkan nilai kategorikal sebagai ordinal atau memiliki hubungan numerik, yang bisa menyebabkan bias dalam model. Metode ini membantu model dalam memahami setiap kategori sebagai entitas terpisah, meningkatkan akurasi prediksi terutama pada algoritma yang sensitif terhadap hubungan numerik, seperti Logistic Regression dan Gradient Boosting.

Metode yang kedua adalah *label encoding* kolom kategorikal yang memiliki urutan inherent, seperti kategori BMI (kurus, normal, overweight, obesitas). Label encoding mengubah setiap kategori menjadi nilai numerik yang berurutan. Label encoding lebih sederhana dan cocok untuk kategori yang memiliki urutan logis, membantu model dalam menangkap informasi ordinal. Metode ini memungkinkan model untuk memahami hubungan ordinal antara kategori, yang bisa meningkatkan kinerja pada algoritma yang memanfaatkan informasi urutan, seperti Decision Tree.

Data yang akan digunakan akan mengalami tahapan pembersihan data, penanganan data yang hilang dan encoding pada variabel kategorikal dilakukan dengan kode yang dapat dilihat pada Gambar 1.

```
[ ] # import library
import pandas as pd

[ ] # Load dataset
df = pd.read_csv("/content/drive/MyDrive/Dataset collection/sleep Quality.csv")
df.drop('Person ID', axis=1, inplace=True)
```

Gambar 1. Proses Persiapan Data

Dataset yang digunakan terdiri dari 13 kolom yaitu kolom *Person ID*, *Gender*, *Age*, *Occupation*, *Sleep Duration*, *Quality of Sleep*, *Physical Activity Level*, *Stress Level*, *BMI Category*, *Blood Pressure*, *Heart Rate*, *Daily Steps*, dan *Sleep Disorder*. Kolom *person ID* tidak berisi informasi yang berharga sehingga akan didrop dari dataset. Proses tersebut dapat dilihat pada Gambar 2.

	gender	Age	Occupation	sleep duration	quality of sleep	Physical activity level	stress level	BMI category	blood pressure	heart rate	Daily steps	sleep disorder
0	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
1	Male	28	Doctor	6.2	6	40	4	Normal	120/80	75	10000	None
2	Male	26	Doctor	6.2	6	40	6	Normal	120/80	75	10000	None
3	Male	26	Sales Representative	5.8	4	30	6	Obese	140/90	88	3000	Sleep Apnea
4	Male	26	Sales Representative	5.8	4	30	6	Obese	140/90	88	3000	Sleep Apnea

Gambar 2. Tampilan Dataset Setelah dilakukan dropping 'Person ID'

Pada kolom '*Blood Pressure*', tekanan darah dibuat dalam format '126/83' yang berarti penulisan antara tekanan darah *Systolic* dan *Diastolic* digabung dengan simbol "/" sebagai delimiter. Hal tersebut akan terbaca sebagai variabel kategori. Maka, kolom tersebut harus diproses untuk memisahkan tekanan systolic dan diastolic menjadi kolom berbeda sehingga data pada kolom tersebut menjadi numerik. Selanjutnya memproses kolom kategori selain kolom '*Sleep Disorder*' menjadi tipe numerik menggunakan *one-hot encoding* dan kolom target yaitu '*Sleep Disorder*' akan diubah menjadi numerik menggunakan *label encoder*. Proses tersebut dapat dilihat pada Gambar 3.

```

] # Preprocess 'Blood Pressure' column
df['Systolic Pressure'] = df['Blood Pressure'].apply(lambda x: int(x.split('/')[0]))
df['Diastolic Pressure'] = df['Blood Pressure'].apply(lambda x: int(x.split('/')[1]))
df.drop(columns=['Blood Pressure'], inplace=True)
    
```

Pemrosesan pemisahan data Systolic dan Diastolic pada kolom 'Blood Pressure'

(a)

Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
M	35	Software Engineer	7.5	4	3	2	Normal	120/80	75	10000	No
F	28	Teacher	8.0	5	2	1	Overweight	130/90	80	8000	No
M	45	Doctor	6.0	2	4	3	Obese	140/100	90	12000	Yes
F	22	Student	9.0	6	1	1	Underweight	110/70	60	15000	No
M	55	Retired	5.0	1	3	4	Obese	150/110	100	9000	Yes

Dataset setelah pemrosesan pada kolom 'Blood Pressure'

(b)

```

# One-hot encoding for categorical columns
data_encoded = pd.get_dummies(df, columns=df.select_dtypes(include='object').columns.drop('Sleep Disorder'))
    
```

Ubah nilai pada kolom yang bertipe katogri menjadi numerik dengan one-hot encoding

(c)

Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder	Occupation_M	Occupation_F	Occupation_Student	Occupation_Retired	Occupation_Teacher
M	35	7.5	4	3	2	Normal	120	75	10000	No	1	0	0	0	0
F	28	8.0	5	2	1	Overweight	130	80	8000	No	0	1	0	0	0
M	45	6.0	2	4	3	Obese	140	90	12000	Yes	0	0	0	0	1
F	22	9.0	6	1	1	Underweight	110	60	15000	No	0	0	1	0	0
M	55	5.0	1	3	4	Obese	150	100	9000	Yes	0	0	0	0	1

Kolom bertipe kategori diubah menjadi numerik menggunakan one-hot encoding

(d)

```

# Label Encoding for target column
from sklearn.preprocessing import LabelEncoder

label_encoder = LabelEncoder()
data_encoded['Sleep Disorder'] = label_encoder.fit_transform(data_encoded['Sleep Disorder'])
    
```

Mengubah isi kolom target "Sleep Disorder" menjadi numerik dengan label encoder

(e)

Gambar 3. Proses persiapan data untuk (a) proses pemisahan data, (b) Dataset yang telah dilakukan pemisahan data, (c), (d) dan (e) perubahan tipe data

2.3 Tahapan Pemilihan Features

Features yang digunakan untuk memprediksi ada atau tidak nya gangguan tidur adalah *Gender, Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Blood Pressure, Heart Rate dan Daily Steps* [1][2][4][6][9][11][13][17][21][22]. Dari *features* tersebut juga akan digali lagi *features* apa saja yang memiliki nilai informasi yang lebih memiliki pengaruh terhadap model prediksi yang dihasilkan. Selanjutnya data set akan dipisah menjadi data *features* (X) dan target (y). Selanjutnya data *features* akan dilakukan *scaling* atau penskalaan. Hal ini dilakukan agar nilai pada *features* memiliki nilai rentang yang serupa atau sebanding. Metode *scaling* yang digunakan pada tulisan ini adalah *MinMaxScaler*. Teknik ini cocok untuk *features* dengan distribusi tidak normal atau ingin menghindari dampak *outlier* [17][18][19][23]. Proses tersebut dapat dilihat pada Gambar 4.

```

# Split data into features (X) and target (y)
X = data_encoded.drop(columns=['Sleep Disorder'])
y = data_encoded['Sleep Disorder']
    
```

Memisahkan data *features* dan data target

(a)

```

# Feature Scaling
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
X_scaled_df = pd.DataFrame(X_scaled, columns=X.columns)
    
```

Scaling pada data *features*

(b)

Gambar 4. Proses Pemilihan *features*

2.4 Pemilihan Model

Model prediksi akan dibangun menggunakan metode *Logistic Regression, Decision Tree* dan *Gradient Boosting*. *Logistic Regression* merupakan model yang sederhana dan mudah diinterpretasikan. Model ini membantu dalam memahami hubungan antara fitur-fitur input dan kemungkinan terjadinya gangguan tidur. *Logistic Regression* sering digunakan sebagai model *baseline* untuk membandingkan performa model yang lebih kompleks. Hal ini memungkinkan untuk menilai apakah model yang lebih kompleks memberikan peningkatan

kinerja yang signifikan. Decision Tree mampu menangani hubungan non-linear antara fitur-fitur. Ini berarti model ini dapat menangkap interaksi kompleks antara berbagai faktor yang mempengaruhi gangguan tidur. Decision Tree menghasilkan model yang mudah diinterpretasikan dalam bentuk diagram pohon, yang memudahkan pemahaman tentang bagaimana keputusan diambil berdasarkan fitur-fitur input. Gradient Boosting adalah teknik ensemble yang sering memberikan kinerja sangat baik dalam berbagai kompetisi machine learning. Teknik ini menggabungkan beberapa pohon keputusan yang lemah menjadi model yang kuat, meningkatkan akurasi prediksi. Gradient Boosting menggunakan metode seperti shrinkage dan subsampling yang membantu dalam mengurangi overfitting, sehingga model dapat menggeneralisasi dengan baik pada data baru. Dari ketiga metode ini nanti nya akan dilihat model mana yang memiliki kinerja yang paling baik. Sebelum melakukan proses training menggunakan ketiga model machine learning yang telah ditentukan. Data features dan target yang dimiliki akan dipisah menjadi data training dan data testing. Karena data pada target tidak seimbang, maka akan digunakan stratify untuk menyeimbangkan data target[2][3][5]. Hasil pengukuran kinerja model logistic regression dapat dilihat pada Gambar 5.

```

evaluation_results = evaluate_model_performance(model_lr, X_test, y_test)
print("Evaluation results for Logistic Regression Model:")
for metric, value in evaluation_results.items():
    print(f"{metric}: {value}")

Evaluation results for Logistic Regression Model:
Accuracy: 0.9333333333333333
Precision: 0.943672514619883
Recall: 0.9333333333333333
F1-score: 0.9351195445920303
AUC-ROC Score: 0.983781058258915

```

Gambar 5. Hasil pengukuran kinerja model *logistic regression*

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Logistic Regression memiliki performa yang baik dalam memprediksi kelas target. Selanjutnya, akan di observasi fitur apa yang memberikan dampak yang signifikan terhadap model prediksi. Observasi fitur ini dapat dilakukan menggunakan features importance, hasil proses tersebut dapat dilihat pada Gambar 6.

```

] #Evaluasi kepentingan fitur
feature_importance = model_lr.coef_[0]

# Cetak kepentingan fitur
for i, feature in enumerate(X_scaled_df.columns):
    print(f"{feature}: {feature_importance[i]}")

Age: 0.5071351772042944
Sleep Duration: -1.206060900176208
Quality of Sleep: -0.4131284961563838
Physical Activity Level: -0.40393874195466817
Stress Level: -0.11009779598274612
Heart Rate: -0.8068250510211334
Daily Steps: -0.26737973784226465
Systolic Pressure: 1.0090533182033865
Diastolic Pressure: 0.821748003428129
Gender_Female: -0.11067573796924567
Gender_Male: 0.111058580827199
Occupation_Accountant: 1.0294416944530016
Occupation_Doctor: -0.2548341980233861
Occupation_Engineer: 0.5756981011075553
Occupation_Lawyer: -0.20685679285130595
Occupation_Manager: -0.1367515327403719
Occupation_Nurse: -1.4596968546804747
Occupation_Sales_Representative: -0.7490107615441272
Occupation_Salesperson: 1.1381499352996172
Occupation_Scientist: -0.6920880789010767
Occupation_Software_Engineer: 0.15237418541258008
Occupation_Teacher: 0.6039571453259442
BMI_Category_Normal: -0.38503928318970343
BMI_Category_Normal_Weight: 0.22904489792041496
BMI_Category_Obese: 0.07487147119560032
BMI_Category_Overweight: 0.08150575693164204

```

Gambar 6. Hasil prediksi dengan model *logistic regression*

Berdasarkan Gambar 6 dapat dilihat features dengan koefisien yang memiliki nilai absolut yang lebih besar cenderung memiliki pengaruh yang lebih signifikan dalam memprediksi gangguan tidur. Features yang memiliki koefisien yang signifikan antara lain: Usia (*Age*) Durasi Tidur (*Sleep Duration*), Kategori BMI (*BMI Category*), Tekanan Darah Systolik (*Systolic_BP*), Tekanan Darah Diastolik (*Diastolic_BP*), Pekerjaan (*Occupation*), seperti *Accountant*, *Nurse*, dan *Salesperson* sehingga dapat disimpulkan bahwa usia, durasi tidur, kategori BMI, tekanan darah, dan pekerjaan (terutama Akuntan, Perawat, dan *Salesperson*). Selanjutnya, akan digunakan *features* yang terpilih berdasarkan *feature importance* sehingga didapatkan data *features* yang baru dan dilakukan pemisahan data yaitu: data training dan data testing pada data *features* yang baru. Selanjutnya model *logistic regression* dilatih menggunakan data *features* yang baru. Selanjutnya model *logistic regression* dilatih menggunakan data *features* yang baru akan diukur kinerjanya menggunakan metrik pengukuran yang sudah ditentukan. Hasil pengukuran dapat dilihat pada Gambar 7.

```
evaluation_results = evaluate_model_performance(model_new, X_test_new, y_test_new)
print("Evaluation results for Logistic Regression Model with new features:")
for metric, value in evaluation_results.items():
    print(f"{metric}: {value}")

Evaluation results for Logistic Regression Model with new features:
Accuracy: 0.9466666666666667
Precision: 0.9533333333333334
Recall: 0.9466666666666667
F1-score: 0.9477789661319873
AUC-ROC Score: 0.9793857529919201
```

Gambar 7. Hasil pengukuran kinerja dengan metrik pengukuran yang sudah ditentukan

Berdasarkan Gambar 7 dapat dilihat hasil pengukuran kinerja berdasarkan metrik akurasi, presisi, recall, F1-score dan AUC-ROC score, dapat disimpulkan bahwa model yang menggunakan *features* terpilih memiliki peningkatan nilai evaluasi secara keseluruhan. Meskipun terjadi peningkatan nilai evaluasi pada model dengan *features* terpilih, namun perbedaan tidak signifikan. Namun terbukti *features* terpilih menyimpan informasi tambahan yang berguna untuk prediksi yang lebih baik. Tetapi metrik AUC-ROC score memiliki sedikit penurunan. Namun hal ini masih positif dan menandakan bahwa model mampu membedakan kelas dengan baik. Selanjutnya dilakukan prediksi gangguan tidur menggunakan model *decision tree* berdasarkan metrik yang ditentukan[7]. Hasil pengukuran dapat dilihat pada Gambar 8.

```
evaluation_results = evaluate_model_performance(model_dt, X_test, y_test)
print("Evaluation results for Decision Tree:")
for metric, value in evaluation_results.items():
    print(f"{metric}: {value}")

Evaluation results for Decision Tree:
Accuracy: 0.9466666666666667
Precision: 0.9524444444444445
Recall: 0.9466666666666667
F1-score: 0.9475296096751462
AUC-ROC Score: 0.9588246765881101
```

Gambar 8. Hasil pengukuran dengan *decision tree*

Gambar 8 merupakan hasil pengukuran kinerja model prediksi menggunakan *decision tree*. pada model di atas semua *features* di dataset dimasukkan. Pada *decision tree*, terdapat beberapa cara untuk menentukan fitur yang paling berpengaruh dalam model *decision tree* yaitu Menggunakan *Recursive Feature Elimination (RFE)*[19][24].

Pada tulisan ini akan digunakan teknik RFE karena teknik ini menghilangkan fitur satu per satu dan membangun model pada fitur yang tersisa untuk mengevaluasi kinerja. Dengan

cara ini, dapat ditentukan *features* yang paling penting dengan melihat perubahan dalam kinerja model saat fitur dihilangkan. Berdasarkan RFE, *features* yang memiliki dampak signifikan terhadap model prediksi gangguan tidur adalah 'Age', 'Sleep Duration', 'Daily Steps', 'Systolic Pressure', 'Gender_Female', 'BMI Category_Overweight'. Selanjutnya akan dibuat Data *features* baru yang hanya berisi *features* yang terpilih dengan proses yang sama dengan sebelumnya. Hasil pengukuran kinerja model tersebut dapat dilihat pada Gambar 9.

```

evaluation_results = evaluate_model_performance(new_model_dt, x_test_RFE, y_test_RFE)
print("Evaluation results for Decision Tree with new features:")
for metric, value in evaluation_results.items():
    print(f"{metric}: {value}")

Evaluation results for Decision Tree with new features:
Accuracy: 0.9466666666666667
Precision: 0.9466666666666667
Recall: 0.9466666666666667
F1-score: 0.9466666666666667
AUC-ROC Score: 0.9687815442561286

```

Gambar 9. Hasil pengukuran dengan model *decision tree* dengan *features* baru yang terpilih

Berdasarkan hasil pengukuran dapat disimpulkan, kedua model *decision tree* yang menggunakan seluruh *features* maupun yang hanya menggunakan *features* berdasarkan selisih RFE memiliki stabilitas kinerja yang cukup stabil. Hal ini terlihat dari metrik evaluasi antara kedua model. Meskipun model *decision tree* yang menggunakan *features* RFE yang terseleksi memiliki nilai evaluasi yang serupa dengan model awal, terdapat peningkatan yang signifikan dalam AUC-ROC score. Hal ini menunjukkan bahwa *features* RFE berhasil meningkatkan kemampuan model dalam membedakan kelas.

Model terakhir yang akan digunakan untuk memprediksi gangguan tidur adalah menggunakan *gradient boosting*. Hasil pengukuran model prediksi gangguan tidur menggunakan *gradient boosting* dapat dilihat pada Gambar 10.

```

evaluation_results = evaluate_model_performance(model_gb, x_test, y_test)
print("Evaluation results for Gradient Bossting:")
for metric, value in evaluation_results.items():
    print(f"{metric}: {value}")

Evaluation results for Gradient Bossting:
Accuracy: 0.96
Precision: 0.962248366013072
Recall: 0.96
F1-score: 0.9601115436905203
AUC-ROC Score: 0.9941732580037664

```

Gambar 10. Hasil pengukuran gangguan tidur menggunakan *gradient boosting*

Selanjutnya akan di observasi feature apa yang memiliki dampak signifikan dari model yang terbentuk pada model *gradient boosting* dapat menggunakan SHAP (*Shapley Additive exPlanations*)[25][26][27][28]. SHAP adalah teknik interpretasi model yang memberikan penjelasan tentang prediksi model dengan menganalisis kontribusi relatif dari setiap fitur terhadap setiap prediksi. Hasil Analisa SHAP dapat dilihat pada Gambar 11.

```

Age: 0.17536434599886264
Sleep Duration: 0.003417357282407777
Quality of Sleep: 0.007838014735230783
Physical Activity Level: 0.0104827056637729
Stress Level: 0.000745224563398173
Heart Rate: 0.0014442748728157977
Daily Steps: 0.004254145721683961
Systolic Pressure: 0.0022488927362183007
Diastolic Pressure: 0.000647953885321464
Gender Female: 0.03746484891357097
Gender Male: 0.016251878176718838
Occupation Accountant: 0.00536795189928372
Occupation Doctor: 0.0013934293615326436
Occupation Engineer: 0.007823678726283917
Occupation Lawyer: 0.0020861403713967586
Occupation Manager: 0.0065317743472245145
Occupation Nurse: 0.007492268827544321
Occupation Sales Representative: 0.004193374481460864
Occupation Salesperson: 0.004245944897671856
Occupation Scientist: 0.007209183818571595
Occupation Software Engineer: 0.006376082309798336
Occupation Teacher: 0.005665717479478248
BMI Category Normal: 0.6491690540132723
BMI Category Normal Weight: 0.008859140077818636
BMI Category Obese: 0.0035326362359498448
BMI Category Overweight: 0.00988299720202621

```

Gambar 11. Hasil Analisa SHAP

Berdasarkan hasil analisa menggunakan SHAP, *features* yang memiliki dampak paling signifikan untuk model prediksi gangguan tidur menggunakan gradient boosting adalah "BMI_Category_Normal" yaitu sebesar 0.649. *Features* "Gender_Female" dan "Gender Male" juga memiliki dampak yang cukup signifikan. Dan juga *features* Occupation juga memiliki pengaruh pada prediksi gangguan tidur khususnya pada *features* "Occupation_Doctor", "Occupation_Engineer", "Occupation_Scientist", dan "Occupation_Software Engineer". *Features* "Stress Level", "Heart Rate", dan "Sleep Duration" memiliki dampak yang relatif kecil terhadap prediksi model. Hal ini bisa jadi karena faktor-faktor tersebut memiliki pengaruh yang lebih kompleks atau tidak linear terhadap kesehatan[1][2][8][13][21][29].

3. HASIL DAN PEMBAHASAN

Receiver Operating Characteristic (ROC) Curve adalah grafik yang menunjukkan kemampuan model untuk membedakan antara kelas positif dan kelas negatif. Grafik ini dibuat dengan plot True Positive Rate (TPR) terhadap False Positive Rate (FPR) pada berbagai threshold. TPR, juga dikenal sebagai Sensitivity atau Recall, adalah rasio data positif yang benar-benar terdeteksi sebagai positif oleh model. Ini menunjukkan seberapa baik model dalam mendeteksi kasus positif. False Positive Rate (FPR) adalah rasio data negatif yang salah terdeteksi sebagai positif oleh model. Ini menunjukkan seberapa sering model salah mendeteksi kasus negatif sebagai positif. Area Under the Curve (AUC) adalah nilai yang mewakili luas di bawah ROC Curve. Nilai AUC berkisar antara 0 dan 1. Semakin tinggi nilai AUC, semakin baik model dalam membedakan antara kelas positif dan negatif. AUC sebesar 1 menunjukkan model sempurna.

Berdasarkan Hasil Pengujian kinerja model yang telah dilakukan dapat dilihat bahwa model prediksi yang dibangun memiliki kinerja yang lebih tinggi dibanding dengan *base* model. Karena model prediksi yang dibangun adalah klasifikasi, maka yang *baseline* yang digunakan adalah proporsi terbesar pada kelas. *Base* model untuk klasifikasi adalah 0.56[19], [24], [27], [30]. Nilai ini akan menjadi dasar untuk mengukur kinerja model prediksi yang dilatih yang dilatih. Hasil pengukuran kinerja model prediksi dengan beberapa model AI dapat dilihat pada Tabel 1.

Tabel 1. Hasil pengukuran kinerja model prediksi

Model Prediksi	Akurasi	Presisi	Recall	F1-Score	AUC-ROC Score
<i>Logistic Regression</i>	0,93	0,94	0,93	0,93	0,98
<i>Logistic Regression with Selected Features</i>	0,94	0,95	0,94	0,94	0,97
<i>Decision Tree</i>	0,94	0,95	0,94	0,94	0,95
<i>Decision Tree with Selected Features</i>	0,94	0,94	0,94	0,94	0,96
<i>Gradient Boosting</i>	0,96	0,96	0,96	0,96	0,99

Berdasarkan Tabel 1 dapat dilihat model prediksi yang dilatih dengan gradient boosting memiliki kinerja yang paling baik dibandingkan dengan model yang lain. Berdasarkan hasil analisis menggunakan tiga pendekatan yang berbeda (*importance_features*, RFE, dan SHAP) terhadap model prediksi gangguan tidur adalah meskipun menggunakan metode analisa yang berbeda, beberapa *features* seperti usia (*ages*), durasi tidur (*sleep duration*), kategori BMI dan pekerjaan memiliki pengaruh yang kuat terhadap prediksi gangguan tidur. *Features BMI Category Normal* memiliki dampak yang sangat signifikan pada prediksi gangguan tidur dengan menggunakan metode *gradient boosting*. Hal ini menunjukkan bahwa individu dengan berat badan normal cenderung memiliki pola tidur yang lebih baik dibandingkan mereka yang memiliki kelebihan berat badan. *Features* terkait dengan pekerjaan seperti *Occupation_Doctor*, *Occupation_Engineer*, *Occupation_Scientist*, dan *Occupation_Software Engineer* juga terbukti signifikan dalam analisis. Ini menunjukkan bahwa faktor-faktor seperti tingkat stres atau jadwal kerja yang tidak teratur yang terkait dengan profesi tertentu dapat memengaruhi pola tidur seseorang. Beberapa fitur seperti *Stress Level*, *Heart Rate*, dan *Sleep Duration* mungkin memiliki dampak yang lebih kompleks atau tidak linear terhadap prediksi gangguan tidur. Hal ini menunjukkan bahwa hubungan antara fitur-fitur ini dan gangguan tidur mungkin lebih kompleks daripada yang diperkirakan, dan memerlukan penelitian lebih lanjut untuk memahaminya dengan lebih baik. Dengan nilai AUC – ROC Score yang tinggi, kita dapat yakin bahwa model ini memiliki kemampuan yang sangat baik untuk mendeteksi gangguan tidur, yang dapat bermanfaat untuk intervensi kesehatan lebih dini dan efektif.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi gangguan tidur dengan akurasi yang sangat tinggi, yaitu 99% berdasarkan AUC – ROC Score. Hasil ini menunjukkan potensi besar untuk aplikasi praktis dalam praktik klinis dan kebijakan kesehatan. Model prediksi ini dapat digunakan oleh profesional kesehatan untuk mendeteksi dini gangguan tidur pada pasien. Dengan identifikasi dini, intervensi yang tepat bisa dilakukan lebih awal, sehingga mengurangi dampak negatif gangguan tidur. Model ini memungkinkan pendekatan yang lebih personal dalam perawatan pasien. Data yang dikumpulkan dari pasien dapat diinput ke dalam model untuk memberikan prediksi yang lebih spesifik dan relevan, sehingga rencana perawatan bisa disesuaikan dengan kebutuhan individu. Model prediksi bisa digunakan untuk memantau pasien yang sudah didiagnosis dengan gangguan tidur, membantu dalam evaluasi efektivitas perawatan dan membuat keputusan berdasarkan data. Hasil penelitian ini dapat digunakan oleh pembuat kebijakan untuk mengembangkan program kesehatan yang lebih

efektif berdasarkan data dan bukti ilmiah. Misalnya, program edukasi tentang pentingnya tidur yang cukup dan sehat. Model ini dapat membantu dalam mengidentifikasi populasi yang berisiko tinggi mengalami gangguan tidur, memungkinkan alokasi sumber daya yang lebih tepat sasaran untuk pencegahan dan perawatan.

Data dari model prediksi ini dapat digunakan untuk mendukung kampanye kesehatan publik, meningkatkan kesadaran masyarakat tentang pentingnya tidur yang sehat dan faktor-faktor yang mempengaruhi kualitas tidur. Banyak penelitian sebelumnya telah mengeksplorasi faktor-faktor individu yang mempengaruhi gangguan tidur. Namun, penelitian ini mengintegrasikan berbagai faktor seperti durasi tidur, kualitas tidur, tingkat aktivitas fisik, tingkat stres, BMI, tekanan darah, detak jantung, dan jumlah langkah harian, memberikan pendekatan yang lebih komprehensif. Dengan menggunakan beberapa algoritma machine learning seperti Logistic Regression, Decision Tree, dan Gradient Boosting, penelitian ini memperluas metodologi yang digunakan dalam penelitian gangguan tidur, menunjukkan keunggulan dari model ensemble dalam meningkatkan akurasi prediksi. Hasil penelitian ini menegaskan pentingnya implementasi teknologi machine learning dalam praktik klinis dan kebijakan kesehatan, memberikan bukti konkret bahwa penggunaan model prediksi berbasis data dapat meningkatkan efektivitas deteksi dan intervensi gangguan tidur. Dengan demikian, penelitian ini tidak hanya menambah wawasan baru dalam literatur yang ada, tetapi juga memberikan alat praktis yang dapat digunakan dalam berbagai setting kesehatan untuk mengatasi masalah gangguan tidur secara lebih efektif.

Pemahaman yang mendalam mengenai faktor-faktor yang berpengaruh terhadap gangguan tidur dapat memberikan pemahaman yang berharga dalam usaha pencegahan dan pengelolaan gangguan tidur. Dengan mempertimbangkan *features* yang memiliki dampak yang signifikan, seperti usia, durasi tidur, kategori BMI, dan pekerjaan, serta memahami kompleksitas hubungan dengan *features* lain seperti tingkat stres dan detak jantung, langkah-langkah intervensi yang lebih sesuai dan efektif dapat diambil. Rencana selanjutnya dapat melibatkan penelitian yang lebih mendalam untuk memahami keterkaitan yang lebih kompleks antara fitur-fitur dan gangguan tidur, serta pengembangan model prediksi yang lebih maju dan akurat untuk mendukung diagnosis dan pengelolaan gangguan tidur dengan lebih efektif. Selain itu, usaha untuk meningkatkan kualitas data dengan mengumpulkan informasi tambahan dan melakukan pemrosesan yang teliti juga dapat meningkatkan akurasi dan utilitas dari model prediksi tersebut.

REFERENSI

- [1] J. Zhang *et al.*, "Automatic detection of obstructive sleep apnea events using a deep CNN-LSTM model," *Comput Intell Neurosci*, vol. 2021, 2021, doi: 10.1155/2021/5594733.
- [2] E. Urtnasan, J. U. Park, E. Y. Joo, and K. J. Lee, "Identification of Sleep Apnea Severity Based on Deep Learning from a Short-term Normal ECG," *J Korean Med Sci*, vol. 35, no. 47, Dec. 2020, doi: 10.3346/jkms.2020.35.e399.
- [3] K. Deep, E. Urtnasan, J.-U. Park, E. Yeon Joo, and K.-J. Lee, "Citation: Convolutional Recurrent Model for Automatic Scoring Sleep Stages Based on Single-Lead ECG Signal. Deep Convolutional Recurrent Model for Automatic Scoring Sleep Stages Based on Single-Lead ECG Signal," *Diagnostics* 2022, vol. 12, p. 1235, 2022, doi: 10.3390/diagnostics.

- [4] Y. K. Wang and C. Y. Chen, "Integrating Mobile Devices and Wearable Technology for Optimal Sleep Conditions," *Applied Sciences (Switzerland)*, vol. 13, no. 17, Sep. 2023, doi: 10.3390/app13179921.
- [5] M. A. Hamza *et al.*, "Wearables-Assisted Smart Health Monitoring for Sleep Quality Prediction Using Optimal Deep Learning," *Sustainability (Switzerland)*, vol. 15, no. 2, Jan. 2023, doi: 10.3390/su15021084.
- [6] A. A. Hidayat, A. Budiarto, and B. Pardamean, "Long Short-Term Memory-based Models for Sleep Quality Prediction from Wearable Device Time Series Data," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 1062–1069. doi: 10.1016/j.procs.2023.10.616.
- [7] M. Mega Santoni, N. Chamidah, and N. Matondang, "Prediction of Hypertension using Decision Tree, Naïve Bayes and Artificial Neural Networks in KNIME Analytics Platform," 2020.
- [8] A. Arora, P. Chakraborty, and M. P. S. Bhatia, "Analysis of Data from Wearable Sensors for Sleep Quality Estimation and Prediction Using Deep Learning," *Arab J Sci Eng*, vol. 45, no. 12, pp. 10793–10812, Dec. 2020, doi: 10.1007/s13369-020-04877-w.
- [9] C. Gao *et al.*, "Sleep Duration/Quality with Health Outcomes: An Umbrella Review of Meta-Analyses of Prospective Studies," Jan. 20, 2022, *Frontiers Media S.A.* doi: 10.3389/fmed.2021.813943.
- [10] A. H. Amalia, "Sistem Pakar Pembantu Diagnosa Penyakit Gangguan Tidur Dengan Metode Certainty Factor," Skripsi, Universitas Muhammadiyah Ponorogo, 2021.
- [11] A. Hajj *et al.*, "Impact of sleep disorders and other factors on the quality of life in general population a cross-sectional study," *Journal of Nervous and Mental Disease*, vol. 207, no. 5, pp. 333–339, May 2019, doi: 10.1097/NMD.0000000000000968.
- [12] S. Suratna and A. Widarma, "Sistem Pakar Diagnosa Penyakit Insomnia Menggunakan Metode Fuzzy Berbasis Web," *Hello World Jurnal Ilmu Komputer*, vol. 1, no. 1, pp. 53–64, May 2022, doi: 10.56211/helloworld.v1i1.12.
- [13] 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.
- [14] W. N. Alfi *et al.*, "Hubungan Kualitas Tidur Dengan Tekanan Darah Pasien Hipertensi Di Puskesmas Mojolangu Kota Malang," *Jurnal Berkala Epidemiologi*, pp. 25–36, 2018, doi: 10.20473/jbe.v6i1.2018.
- [15] R. Fernando and R. Hidayat, "Hubungan Lama Penggunaan Media Sosial Dengan Kejadian Insomnia Pada Mahasiswa Fakultas Ilmu Kesehatan Universitas Pahlawan Tuanku Tambusai Tahun 2020," *Jurnal Ners*, vol. 4, no. 2, pp. 83–89, 2020.
- [16] N. A. ZALUKHU, "Hubungan Kualitas Tidur Dengan Kadar HBA1C Pada Penderita Diabetes Melitus Tipe 2 Di RSUD Dr. Pirngadi Medan Tahun 2019," Skripsi, Universitas HKBP Nommensen, 2020.
- [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, Jan. 2023, doi: 10.2196/46520.
- [18] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *Journal of Applied Computer Science and Technology (JACOST)*, vol. 4, no. 1, pp. 2723–1453, 2023, doi: 10.52158/jacost.491.

- [19] A. Satria, R. M. Badri, and I. Safitri, "Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning," *Digital Transformation Technology*, vol. 3, no. 2, pp. 389–398, Sep. 2023, doi: 10.47709/digitech.v3i2.2852.
- [20] "Dataset." Accessed: Jan. 02, 2024. [Online]. Available: <https://www.kaggle.com/datasets/henryshan/sleep-health-and-lifestyle>.
- [21] C. Sun, S. Hong, J. Wang, X. Dong, F. Han, and H. Li, "A systematic review of deep learning methods for modelling electrocardiograms during sleep," Aug. 31, 2022, *Institute of Physics*. doi: 10.1088/1361-6579/ac826e.
- [22] M. A. Albqoor and A. M. Shaheen, "Sleep quality, sleep latency, and sleep duration: a national comparative study of university students in Jordan," *Springer Nature Switzerland*, pp. 1–8, 2021.
- [23] C. Gao *et al.*, "Sleep Duration/Quality with Health Outcomes: An Umbrella Review of Meta-Analyses of Prospective Studies," Jan. 20, 2022, *Frontiers Media S.A.* doi: 10.3389/fmed.2021.813943.
- [24] Z. Zeng, W. Peng, and B. Zhao, "Improving the Accuracy of Network Intrusion Detection with Causal Machine Learning," *Security and Communication Networks*, vol. 2021, 2021, doi: 10.1155/2021/8986243.
- [25] I. Wardhana, M. Ariawijaya, V. A. Isnaini, and R. P. Wirman, "Gradient Boosting Machine, Random Forest dan Light GBM untuk Klasifikasi Kacang Kering," *Jurnal Resti*, vol. 6, no. 1, pp. 92–99, 2021.
- [26] D.- Andriansyah and Eka Wulansari Fridayanthie, "Optimization of Support Vector Machine and XGBoost Methods Using Feature Selection to Improve Classification Performance," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 6, no. 2, pp. 484–493, Jan. 2023, doi: 10.31289/jite.v6i2.8373.
- [27] J. Homepage *et al.*, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison of K-Means and K-Medoids on Poor Data Clustering in Indonesia Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," vol. 2, no. 2, pp. 35–41, 2022.
- [28] Y. H. Kao and B. Van Roy, "Learning a factor model via regularized PCA," *Mach Learn*, vol. 91, no. 3, pp. 279–303, 2013, doi: 10.1007/s10994-013-5345-8.
- [29] M. A. Albqoor and A. M. Shaheen, "Sleep quality, sleep latency, and sleep duration: a national comparative study of university students in Jordan," *Sleep and Breathing*, vol. 25, no. 2, pp. 1147–1154, Jun. 2021, doi: 10.1007/s11325-020-02188-w.
- [30] Z. Zhou, Y. Liu, H. Yu, and L. Ren, "The influence of machine learning-based knowledge management model on enterprise organizational capability innovation and industrial development," *PLoS One*, vol. 15, no. 12 December, Dec. 2020, doi: 10.1371/journal.pone.0242253.