PROCEEDING NATIONAL CONFERENCE OF RESEARCH AND COMMUNITY SERVICE SISI INDONESIA
2025
https://conference.sinesia.id/ncrcs-sinesia

Analisis Perbandingan Algoritma *Decision Tree, Random Forest,* Dan *Naïve Bayes* Dalam Klasifikasi Gangguan Tidur

Indah Novita Sari[1], Lukman Fakih Lidimilah, M.Kom[2] , Ahmad Lutfi, M.Kom [3]
[1]indahvitasari943@gmail.com, [2]luky.lukman7@gmail.com, [3] ahmadlutfi.14@gmail.com
[1]Jurusan Sistem Informasi, Sains dan Teknologi, Universitas Ibrahimy
[2]Jurusan Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy
[3]Jurusan Sistem Informasi, Sains dan Teknologi , Universitas Ibrahimy

Info Article
|Submitted: 20 May 2025 **|Revised**: 7 June 2025 **|Accepted**: 9 June 2025
How to cite: Indah Novita Sari, etc., "Analisis Perbandingan Algoritma *Decision Tree, Random Forest*, Dan *Naïve Bayes* Dalam Klasifikasi Gangguan Tidur", *Proceeding National Conference Of Research And Community Service Sisi Indonesia*, 2025, P. 192-209.

ABSTRACT

Data analysis has many methods, one of which is data mining with machine learning-based classification techniques. This study compares three data mining algorithms, namely the Decision Tree algorithm, the Random Forest algorithm, and the Naïve Bayes algorithm. These three algorithms were chosen to determine which algorithm is effective and has the best performance in the classification process. Classification is the process of analyzing data by grouping data samples based on predetermined categories or labels using supervised learning techniques. This study uses sleep disorder cases as its data samples. Sleep disorders are conditions in which a person experiences problems related to the duration, quality or pattern of sleep. Most people ignore the signs of sleep disorders they experience so they do not realize that they could be experiencing sleep disorders. Early detection of sleep disorders can help prevent various serious health problems. This study was conducted to analyze the performance of each algorithm and provide recommendations for the best algorithm for classifying sleep disorders and developing a sleep disorder detection system. This study compares the performance results of the Decision Tree algorithm, the Random Forest algorithm, and the Naïve Bayes algorithm. The data used is the Sleep Health and Lifestyle Dataset obtained online through Open Data Source, namely from the Kaggle Platform, which consists of 374 data and 13 attributes. Based on the test results of Decision Tree, Random Forest, and Naïve Bayes, the average accuracy of Decision Tree was 91.49%, Random Forest 87.23%, and Naïve Bayes 76.59%. It can be concluded that the Decision Tree algorithm provides the best performance in classifying sleep disorders.

Keyword: *Classification, Sleep Disorders, Decision Trees, Random Forest, Naïve Bayes*

ABSTRAK

Analisis data memiliki banyak metode salah satunya adalah data mining dengan teknik klasifikasi berbasis machine learning. Penelitian ini membandingkan tiga algoritma data mining yaitu algoritma *Decision Tree,* algoritma *Random Forest*, dan algoritma *Naïve Bayes*. Ketiga algoritma ini dipilih untuk menentukan algoritma mana yang efektif dan memiliki performa terbaik dalam proses klasifikasi. Klasifikasi merupakan proses menganalisis data dengan mengelompokkan sampel data berdasarkan kategori atau label yang sudah ditentukan sebelumnya menggunakan teknik *supervised learning*. Penelitian ini menggunakan kasus gangguan tidur sebagai sample datanya. Gangguan tidur adalah kondisi seseorang yang mengalami masalah terkait durasi, kualitas atau pola tidur. Kebanyakan orang tidak menghiraukan tanda – tanda gangguan tidur yang mereka alami sehingga mereka tidak menyadari bahwa bisa saja mereka mengalami gangguan tidur. Deteksi dini gangguan tidur dapat membantu mencegah berbagai masalah kesehatan yang serius. Penelitian ini dilakukan untuk menganalisis performa dari masing – masing algoritma dan memberikan rekomendasi algoritma terbaik untuk klasifikasi gangguan tidur serta pengembangan sistem deteksi gangguan tidur. Penelitian ini membandingkan hasil performa dari algoritma *Decision Tree,* algoritma *Random Forest*, dan algoritma *Naïve Bayes*. Data yang digunakan yaitu *Sleep Health*

This work is licensed under CC BY-SA 4.0

and Lifestyle Dataset yang diperoleh secara online melalui *Open Data Source* yakni dari *Platform Kaggle* yang terdiri dari 374 data dan 13 atribut. Berdasarkan hasil pengujian *Decision Tree, Random Forest*, dan *Naïve Bayes*, diperoleh hasil rata – rata akurasi *Decision Tree* 91.49%, *Random Forest* 87.23%, dan *Naïve Bayes* 76.59%. Dapat disimpulkan bahwa algoritma *Decision Tree* memberikan performa terbaik dalam klasifikasi gangguan tidur.
Kata Kunci: *Klasifikasi, Gangguan Tidur, Decision Tree, Random Forest, Naïve Bayes*

Pendahuluan

Penggunaan data dan teknologi pada zaman yang serba digital seperti sekarang memiliki peran yang penting di berbagai bidang, termasuk juga pada bidang kesehatan(Setyawan, 2016). Berbagai metode bisa digunakan dalam menganalisis data, salah satunya adalah *data mining* dengan menggunakan teknik klasifikasi berbasis *machine learning*(Baharuddin & Tjahyanto, 2022). Dalam *data mining* ada berbagai algoritma yang dapat diterapkan, pada penelitian ini *data mining* yang digunakan yaitu perbandingan dari tiga algoritma *machine learning* yaitu algoritma *Decision Tree,* algoritma *Random Forest,* dan algoritma *Naïve Bayes*. Guna menentukan algoritma mana yang terbukti efektif dan memiliki performa terbaik dalam melakukan klasifikasi maka digunakanlah ketiga algoritma tersebut.

Decision Tree adalah sebuah metode klasifikasi yang berbentuk pohon keputusan yang membuat interpretasi menjadi lebih mudah, perhitungan yang lebih efisien, serta mampu menangani data dengan atribut numerik dan kategorikal. Algoritma *Random Forest*, merupakan pengembangan dari *Decision Tree*, yang mengombinasikan beberapa pohon keputusan yang bekerja secara acak membentuk beberapa pohon keputusan, lalu hasil prediksi digabungkan dari tiap pohon untuk mendapatkan hasil akhir yang stabil dan akurat. Sementara itu, algoritma *Naive Bayes* merupakan algoritma berbasis *probabilitas* yang dikenal efisien untuk klasifikasi teks dan aplikasi yang melibatkan dataset besar karena memiliki waktu komputasi yang cepat dan relatif sederhana dalam implementasi (Renganathan, 2021).

Banyak penelitian yang menerapkan teknik klasifikasi(Prabowo et al., 2023). Metode klasifikasi menggunakan teknik *supervised learning* dalam mengidentifikasi pola atau hubungan antara data input dan target yang akan di prediksi (Fuji Astari et al., 2024). Algoritma yang digunakan dalam klasifikasi juga berbagai macam dengan cara kerja dari masing masing algoritma juga berbeda tergantung pada karakteristik data yang digunakan(Hassan et al., 2018). Maka dari itu, diperlukan analisis lebih lanjut untuk memastikan keakuratan dari algoritma yang diterapkan dalam penelitian ini.

Penelitian oleh Muhammad Yasir dan Robertus Suraji menggunakan metode *CRISP-DM*. Hasilnya menunjukkan akurasi yang didapatkan algoritma *Naive Bayes* sebesar 90%, untuk algoritma *Decision Tree* 83%, dan untuk algoritma *Random Forest* 87%(Yasir & Suraji, 2023). Penelitian yang dilakukan oleh Yulrio

Brianorman dan Sucipto, algoritma *Random Forest* menunjukkan performa yang lebih unggul dengan akurasi yang tinggi sebesar 59,2%, diikuti oleh algoritma *Decision Tree* dengan akurasi 59,1%, dan algoritma *Naive Bayes* dengan akurasi 58,1%. *Random Forest* terbukti lebih efektif, dengan perkiraan potensi pendaftaran calon mahasiswa berkisar antara 72% hingga 78% dari jumlah total(Brianorman & Sucipto, 2024).

Dalam penelitian ini, sampel data yang akan dianalisis adalah kasus gangguan tidur. Berdasarkan data *National Institute of Neurological Disorders and Stroke*, tidur adalah kondisi di mana aktivitas mental dan fisik melambat, kesadaran berkurang, dan interaksi dengan lingkungan menjadi minim. Tidur sangat penting bagi kesehatan tubuh serta mental, mempengaruhi berbagai sistem tubuh seperti metabolisme, kekebalan, dan suasana hati (Stroke, n.d.). Manusia menghabiskan sekitar 25% hingga 33% dari kehidupannya untuk tidur (Dr. H. Fuad Nashori, S.Psi., M.Si., M.Ag. & Etik Dwi Wulandari, S.Psi., M.Psi., 2017).

Meskipun telah menunjukkan gejala gangguan tidur pada diri mereka, masih banyak orang yang belum menyadari bahwa sebenarnya mereka mengalami gangguan tidur. Gangguan tidur merupakan kondisi yang mencakup masalah pada durasi, kualitas, atau pola tidur seseorang (Kesehatan, 2022). Jenis gangguan tidur yang menjadi tolak ukur pada penelitian ini yaitu tanpa gangguan tidur, *sleep apnea,* dan *insomnia*. *Sleep apnea* adalah kondisi dimana pernapasan terhenti secara berulang saat tidur (Sari, 2024), sedangkan *insomnia* adalah situasi dimana seseorang mengalami sulit untuk tidur atau sering terbangun dalam tidurnya (Mais et al., 2020).

Untuk mencegah terjadinya masalah kesehatan yang serius di kemudian hari, deteksi dini gangguan tidur sangat penting untuk dilakukan. Tiga algoritma tersebut dipilih untuk mengetahui algoritma mana yang paling efektif dan memiliki performa terbaik dalam mengklasifikasikan gangguan tidur, sekaligus mendukung sistem deteksi gangguan tidur sehingga dapat diimplementasikan secara praktis dalam lingkungan operasional, seperti pada fasilitas kesehatan atau aplikasi berbasis teknologi.

Metode Penelitian

Metode dan Tahapan Penelitian

Penelitian ini memanfaatkan teknik analisis data dengan metode klasifikasi guna mengenali jenis gangguan tidur berdasarkan informasi tentang pola hidup dan kondisi kesehatan seseorang. Tahapan penelitian diawali dengan pengumpulan data dari dataset *Sleep Health and Lifestyle* yang diunduh dari platform *Kaggle*. Setelah data terkumpul, tahap pembersihan (*preprocessing*) dilakukan untuk menangani data kosong dan menghilangkan data duplikat. Kemudian, atribut – atribut yang dianggap relevan dipilih melalui proses seleksi,

yang dilanjutkan dengan pengubahan data untuk mengubah kategori menjadi format angka agar dapat dibaca dan diproses oleh algoritma. Selanjutnya, penerapan tiga metode klasifikasi yaitu algoritma *Decision Tree,* algoritma *Random Forest*, dan algoritma *Naive Bayes* untuk membangun model prediksi. Terakhir, kinerja tiap model dianalisis memakai ukuran evaluasi seperti *accuracy, precision*, *recall*, dan *f1-score* guna mengukur tingkat keakuratan model dalam mengklasifikasikan gangguan tidur.

Accuracy adalah ukuran yang menunjukkan seberapa besar tingkat ketepatan model dalam membuat prediksi yang akurat, baik untuk data yang positif maupun negatif, jika dibandingkan dengan jumlah keseluruhan data yang diuji. *Precision* menggambarkan seberapa akurat model dalam mengenali data sebagai data positif, yaitu seberapa besar dari semua data yang diperkirakan positif benar-benar memang positif. *Recall* menunjukkan kemampuan sistem dalam mendeteksi atau mengenali sebanyak mungkin data yang memang benar-benar positif(Azhari et al., 2021).

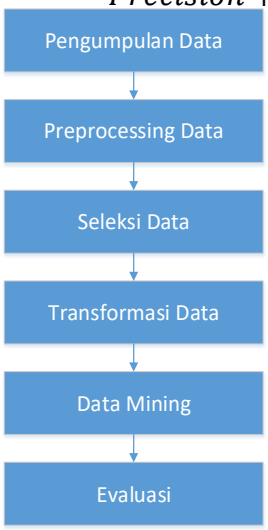
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

F1-Score adalah ukuran yang menggabungkan nilai *precision* dan *recall.* Cara kerjanya adalah dengan mencari nilai Tengah atau rata-rata dari kedua metrik tersebut. Hal ini berguna untuk membantu menilai apakah model bekerja dengan seimbang dalam hal ketepatan dan kemampuan menemukan data positif(Cardova & Hermawan, 2023). Rumus perhitungan untuk *F1-score* adalah sebagai berikut :

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$



Gambar 1. Metode dan Tahapan Penelitian

Jenis Penelitian

Metode kuantitatif digunakan dalam penelitian ini untuk mengumpulkan data, dan menggunakan *Library Online Research* atau penelitian kepustakaan online(Kamaliyah et al., 2023) untuk mengumpulkan informasi yang berkaitan dengan topik penelitian. Hal ini dilakukan melalui kegiatan membaca, mempelajari, dan menganalisis materi untuk mengidentifikasi pokok permasalahan yang didapat dari berbagai sumber di perpustakaan, seperti buku, jurnal, artikel, dan dokumentasi(Zed, 2008).

Metode Pengumpulan Data

Data sekunder yang dipakai dalam penelitian ini didapatkan secara online melalui sumber data terbuka (*Open Data Source*) yakni *Platform Kaggle*. *Kaggle* sendiri merupakan yang menghimpun berbagai informasi dari banyak orang diseluruh dunia, serta menyediakan berbagai dataset yang dapat diolah menggunakan teknik *machine learning* (Rahmat et al., 2023). Dataset yang digunakan yakni *Sleep Health and Lifestyle Dataset* dapat diakses melalui link berikut : www.kaggle.com. Dataset tersebut berisi 374 data mengenai kesehatan tidur dan gaya hidup individu, dengan 13 atribut utama. Dataset ini tidak menyertakan secara langsung informasi spesifik mengenai letak geografis dari masing – masing responden. Data ini disusun serta dipublikasikan oleh seorang kontributor yang bernama Laksika Tharmalingam, yang berasal dari Kilinochchi, Northern Province, Sri Lanka.

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
2	1	Male	27	Software Engineer	06.01	6	42	6	Overweight	126/83	77	4200	None
3	2	Male	28	Doctor	06.02	6	60	8	Normal	125/80	75	10000	None
4	3	Male	28	Doctor	06.02	6	60	8	Normal	125/80	75	10000	None
5	4	Male	28	Sales Representative	05.09	4	30	8	Obese	140/90	85	3000	Sleep Apnea
6	5	Male	28	Sales Representative	05.09	4	30	8	Obese	140/90	85	3000	Sleep Apnea
7	6	Male	28	Software Engineer	05.09	4	30	8	Obese	140/90	85	3000	Insomnia
8	7	Male	29	Teacher	06.03	6	40	7	Obese	140/90	82	3500	Insomnia
9	8	Male	29	Doctor	07.08	7	75	6	Normal	120/80	70	8000	None
10	9	Male	29	Doctor	07.08	7	75	6	Normal	120/80	70	8000	None
11	10	Male	29	Doctor	07.08	7	75	6	Normal	120/80	70	8000	None
12	11	Male	29	Doctor	06.01	6	30	8	Normal	120/80	70	8000	None
13	12	Male	29	Doctor	07.08	7	75	6	Normal	120/80	70	8000	None
14	13	Male	29	Doctor	06.01	6	30	8	Normal	120/80	70	8000	None
15	14	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
16	15	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
17	16	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
18	17	Female	29	Nurse	06.05	5	40	7	ormal Weigh	132/87	80	4000	Sleep Apnea
19	18	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	Sleep Apnea
20	19	Female	29	Nurse	06.05	5	40	7	ormal Weigh	132/87	80	4000	Insomnia
21	20	Male	30	Doctor	07.06	7	75	6	Normal	120/80	70	8000	None
22	21	Male	30	Doctor	07.07	7	75	6	Normal	120/80	70	8000	None
23	22	Male	30	Doctor	07.07	7	75	6	Normal	120/80	70	8000	None
24	23	Male	30	Doctor	07.07	7	75	6	Normal	120/80	70	8000	None

Gambar 2. *Sleep Health and Lifestyle Dataset*

Hasil dan pembahasan

1.1 Pengumpulan Data

Data yang digunakan berjudul "*Sleep Health and Lifestyle Dataset*". Dataset tersebut mencakup informasi mengenai kesehatan dan pola hidup dari 374 individu. Di dalamnya terdapat data tentang usia, jenis kelamin, durasi tidur, tingkat aktivitas fisik, tekanan darah, dan masalah tidur. File data ini berformat CSV dan diunggah ke *Google Colab* untuk dianalisis lebih lanjut. Dataset *Sleep Health and Lifestyle* dapat diunduh langsung dari *Kaggle* menggunakan perintah berikut:

Segmen Program 1. Pengumpulan Data

```
!kaggle datasets download -d uom190346a/sleep-health-and-

lifestyle-dataset

!unzip sleep-health-and-lifestyle-dataset.zip
```

Dataset ini menyajikan sejumlah atribut terkait pola tidur, gaya hidup, dan kondisi Kesehatan seseorang. Di dalamnya mencakup informasi seperti *Gender, Age, Occupation, Sleep Duration* (durasi tidur), *Quality of Sleep* (skor kualitas tidur), *Physical Activity Level* (aktivitas fisik), *Stress Level, BMI Category, Blood Pressure* (tekanan darah), *Heart Rate, Daily Steps, dan Sleep Disorder* (gangguan tidur).

1.2 Preprocessing Data

Tahap *preprocessing* dilakukan guna membersihkan, mentransformasi, dan mempersiapkan data sehingga layak digunakan dalam proses data mining. Langkah ini melibatkan penanganan nilai yang hilang, penghapusan data yang duplikat, normalisasi nilai data, serta konversi variabel kategorikal ke dalam format numerik agar dapat diproses oleh algoritma data mining(Agung et al., 2023).

a. Membersihkan data kosong dan menghapus baris dengan nilai *Sleep Disorder* yang kosong dengan kode berikut :

Segmen Program 2. Menghapus Nilai Kosong

```
df = pd.read_csv("Sleep_health_and_lifestyle_dataset.csv")

df = df.dropna(subset=['Sleep Disorder'])
```

Hasil dari penghapusan nilai kosong pada kolom target *Sleep Disorder* adalah sebagai berikut :

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	NaN
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea

Next steps: Generate code with df | View recommended plots | New interactive sheet

Gambar 3. Hasil Penghapusan Nilai Kosong

b. Mengisi *missing value* dengan kode sebagai berikut :

Segmen Program 3. Mengisi *Missing Value*

```
kolom_numerik = df.select_dtypes(include=['int64',
'float64']).columns
kolom_kategorikal = df.select_dtypes(include=['object']).columns
for kolom in kolom_numerik:
    if df[kolom].isnull().sum() > 0:
        df[kolom] = df[kolom].fillna(df[kolom].mean())
for kolom in kolom_kategorikal:
    if df[kolom].isnull().sum() > 0:
        df[kolom] = df[kolom].fillna(df[kolom].mode()[0])
```

Hasil dari proses mengisi *missing value* adalah sebagai berikut :

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

dtype: int64

c. *Encoding* Kolom Kategorikal dengan kode sebagai berikut :

Segmen Program 4. *Encoding* Kolom Kategorikal

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
for kolom in kolom_kategorikal:
    df[kolom] = le.fit_transform(df[kolom])
```

Dari proses *encoding* pada setiap kolom kategorikal, hasil yang didapatkan adalah sebagai berikut :

Tabel 1. Hasil *Encoding* Kolom Kategorikal

	Gender	*Occupation*	*BMI Category*	*Sleep Disorder*
3	1	5	2	1
4	1	5	2	1

5	1	8	2	0
6	1	9	2	0
16	0	4	1	1

1.3 Seleksi dan Transformasi Data

Dalam tahap seleksi dan transformasi data, kolom *Blood Pressure* dipisahkan menjadi dua bagian, yaitu *Systolic* dan *Diastolic*. Pemisahan ini dilakukan agar data lebih mudah dipahami dan dianalisis, sehingga setiap komponen tekanan darah bisa diteliti secara lebih rinci. Dengan cara ini, hasil analisis menjadi lebih akurat dan bermanfaat untuk langkah-langkah berikutnya.

Segmen Program 5. Seleksi dan Transformasi Data

```
df[['Systolic', 'Diastolic']] = df['Blood

Pressure'].str.split('/', expand=True).astype(int)

df.drop(['Blood Pressure', 'Person ID'], axis=1,

inplace=True)
```

Pada proses seleksi dan transfomasi data menunjukkan hasil sebagai berikut :

Gender	int64
Age	int64
Occupation	int64
Sleep Duration	float64
Quality of Sleep	int64
Physical Activity Level	int64
Stress Level	int64
BMI Category	int64
Heart Rate	int64
Daily Steps	int64
Systolic	int64
Diastolic	int64

dtype: *object*

Proses pembagian data akan membagi dataset menjadi dua bagian utama yaitu data pelatihan (*training*) dan data pengujian (*testing*).

Segmen Program 6. Pembagian Data

```
X = df.drop('Sleep Disorder', axis=1)

y = df['Sleep Disorder']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y,

test_size=0.3, random_state=42)
```

Hasil pembagian data menunjukkan bahwa Ukuran *X_train*: (108, 12), ukuran *X_test*: (47, 12), ukuran *y_train*: (108,), dan ukuran *y_test*: (47,).

1.4 Data Mining dan Evaluasi

Penelitian ini menerapkan tiga algoritma klasifikasi, yaitu algoritma *Decision Tree,* algoritma *Random Forest*, dan algoritma *Naive Bayes*. Ketiga model tersebut dilatih menggunakan data pelatihan, lalu diuji dengan data pengujian. Tujuan dari tahap ini adalah untuk membandingkan seberapa akurat masing-masing model dalam memprediksi tipe gangguan tidur.

a. *Decision Tree*

Segmen Program 7. *Decision Tree*

```
from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import accuracy_score,

classification_report

model = DecisionTreeClassifier(random_state=42)

model.fit(X_train, y_train)

y_pred = model.predict(X_test)

print("Akurasi:", accuracy_score(y_test, y_pred))

print("\nClassification Report:")

print(classification_report(y_test, y_pred))
```

Dari pengujian algoritma *Decision Tree* adalah sebagai berikut :

```
=== Decision Tree ===
Accuracy: 0.9148936170212766
              precision    recall  f1-score   support

           0       0.86      1.00      0.93        25
           1       1.00      0.82      0.90        22

    accuracy                           0.91        47
   macro avg       0.93      0.91      0.91        47
weighted avg       0.93      0.91      0.91        47
```

Gambar 4. Output Klasifikasi *Decision Tree*

Hasil klasifikasi menunjukkan terdapat 25 data termasuk dalam kategori tidak mengalami gangguan tidur (kelas 0) yang menghasilkan nilai *True Negative (TN)* sebesar 25. Untuk data yang termasuk dalam kategori mengalami gangguan tidur (kelas 1) terdapat 22 data, namun hanya 18 data yang dapat terdeteksi dengan benar menjadi nilai *True Positive (TP)*,

sedangkan 4 data lainnya diprediksi tidak mengalami gangguan tidur yakni *False Negative (FN)*. Untuk nilai *False Positive (FP)* adalah 0, karena tidak ditemukan kesalahan dalam klasifikasi.

Tabel 2. ***Confusion Matrix Decision Tree***

	Prediksi = 0	**Prediksi = 1**
Aktual = 0 (negatif)	**25** (TN)	**0** (FP)
Aktual = 1 (positif)	**4** (FN)	**18** (TP)

Untuk perhitungan menggunakan rumus pada algoritma *Decision Tree* adalah sebagai berikut :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{18 + 25}{18 + 25 + 0 + 4} = \frac{43}{47} \approx 0.9149$$

$$Precision = \frac{TP}{TP + FP} = \frac{18}{18 + 0} = \frac{18}{18} \approx 1.00$$

$$Recall = \frac{TP}{TP + FN} = \frac{18}{18 + 4} = \frac{18}{22} \approx 0.8182$$

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{1.00 \cdot 0.8182}{1.00 + 0.8182} = \frac{1.6364}{1.8182} \approx 0.900$$

b. *Random Forest*

Segmen Program 8. ***Random Forest***

```
from sklearn.ensemble import RandomForestClassifier

rf_model = RandomForestClassifier(random_state=42)

rf_model.fit(X_train, y_train)

rf_pred = rf_model.predict(X_test)

from sklearn.metrics import classification_report

print(classification_report(y_test, rf_pred))
```

Hasil yang didapat dari pengujian algoritma *Random Forest* seperti pada gambar berikut :

```
=== Random Forest ===
Accuracy: 0.8723404255319149
              precision    recall  f1-score   support

           0       0.85      0.92      0.88        25
           1       0.90      0.82      0.86        22

    accuracy                           0.87        47
   macro avg       0.88      0.87      0.87        47
weighted avg       0.87      0.87      0.87        47
```

Gambar 5. Output Klasifikasi *Random Forest*

Berdasarkan output yang telah dihasilkan diatas, selanjutnya kita dapat menyusun *confusion matrix.*

1. Untuk kelas 0 :
   - *Support* atau jumlah data sebenarnya untuk kelas 0 = 25
   - *Recall* (0) = 0.92, yang artinya model berhasil mengenali 92% dari data kelas 0.
     Maka :
     $$True\ Negative\ (TN) = 0.92\ \times 25 = 23$$
     Ada 23 data pada kelas 0 yang diprediksi benar.
     $$False\ Negative\ (FN) = 25 - 23 = 2$$
     Pada kelas 0 ada 2 data diprediksi salah sebagai kelas 1.
   - *Precision* kelas 0 = 0.85, dari semua prediksi kelas 0, 85% adalah benar.
     Maka :
     $$\frac{TN}{TN + FP} = \frac{23}{23 + FP} = 0.85 \approx 4$$
     Sekitar 4 data dalam kelas 1 diprediksi sebagai kelas 0.
2. Untuk kelas 1 :
   - *Support* kelas 1 = 22
   - *Recall* = 0.82
     $$True\ Positive\ (TP) = 0.82 - 22 = 18$$
     $$False\ Negative\ (FN) = 22 - 18 = 4$$
   - *Precision* = 0.90
     $$\frac{TP}{TP + FP} = \frac{18}{18 + FP} = 0.90 \approx FP = 2$$

Adapun perhitungan algoritma *Random Forest* menggunakan rumus sebagai berikut :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{18 + 23}{47} = \frac{41}{47} \approx 0.8723$$

$$Precision = \frac{TP}{TP + FP} = \frac{18}{18 + 2} = \frac{18}{20} \approx 0.90$$

$$Recall = \frac{TP}{TP + FN} = \frac{18}{18 + 4} = \frac{18}{22} \approx 0.8182$$

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{0.90 \cdot 0.8182}{0.90 + 0.8182} = \frac{1.4724}{1.718} \approx 0.857$$

c. Naïve Bayes

Segmen Program 9. Naïve Bayes

```
from sklearn.naive_bayes import GaussianNB

nb_model = GaussianNB()

nb_model.fit(X_train, y_train)
```

```
nb_pred = nb_model.predict(X_test)

from sklearn.metrics import classification_report

print(classification_report(y_test, nb_pred))
```

Hasilnya menujukkan :

```
=== Naive Bayes ===
Accuracy: 0.7659574468085106
              precision    recall  f1-score   support

           0       0.77      0.80      0.78        25
           1       0.76      0.73      0.74        22

    accuracy                           0.77        47
   macro avg       0.77      0.76      0.76        47
weighted avg       0.77      0.77      0.77        47
```

Gambar 6. Output Klasifikasi *Naïve Bayes*

Dari hasil tersebut maka *confusion matrix* yang diperoleh yaitu :

Tabel 3. *Confusion Matrix Naïve Bayes*

	Predicted = 0	**Predicted = 1**
Actual = 0	TN = 20	FP = 5
Actual = 1	FN = 6	TP = 16

Rumus dasar perhitungan *accuracy, precision, recall,* dan *f1-score* pada algoritma *Naïve Bayes* sebagai berikut :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{16 + 20}{16 + 20 + 5 + 6} = \frac{36}{47} \approx 0.766$$

$$Precision = \frac{TP}{TP + FP} = \frac{16}{16 + 5} = \frac{16}{2} \approx 0.762$$

$$Recall = \frac{TP}{TP + FN} = \frac{16}{16 + 6} = \frac{16}{22} \approx 0.727$$

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{0.762 \cdot 0.727}{0.762 + 0.727} = 2 \cdot \frac{0.554}{1.489} \approx 0.744$$

1.5 Visualisasi Hasil dan Perbandingan

Segmen Program 10. Visualisasi Hasil dan Perbandingan

```
import matplotlib.pyplot as plt

import seaborn as sns

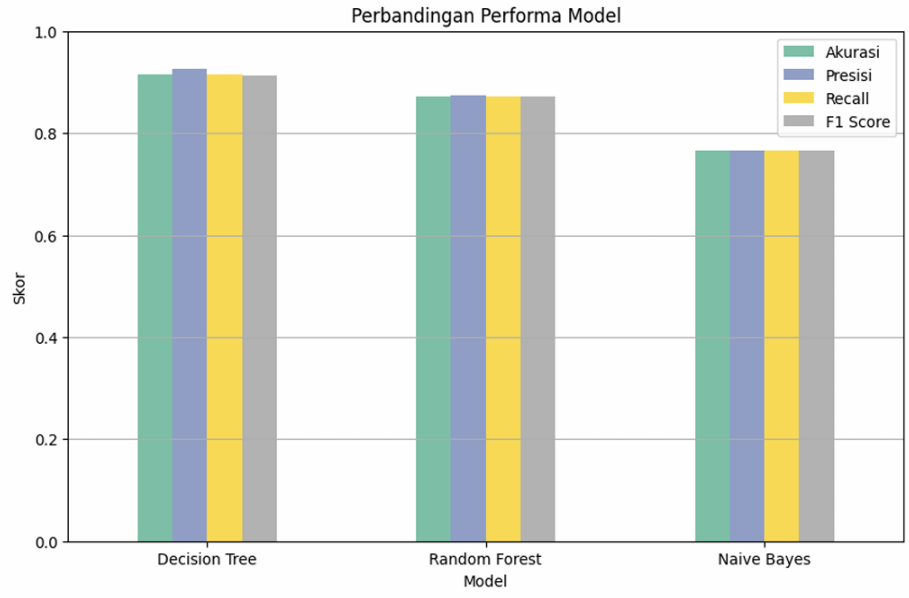
comparison = pd.DataFrame({
```

```
    "Model": ["Decision Tree", "Random Forest", "Naive
Bayes"],
    "Akurasi": [
        accuracy_score(y_test, dt_pred),
        accuracy_score(y_test, rf_pred),
        accuracy_score(y_test, nb_pred)
    ],
    "Presisi": [
        precision_score(y_test, dt_pred,
average='weighted', zero_division=0),
        precision_score(y_test, rf_pred,
average='weighted', zero_division=0),
        precision_score(y_test, nb_pred,
average='weighted', zero_division=0)
    ],
    "Recall": [
        recall_score(y_test, dt_pred, average='weighted',
zero_division=0),
        recall_score(y_test, rf_pred, average='weighted',
zero_division=0),
        recall_score(y_test, nb_pred, average='weighted',
zero_division=0)
    ],
    "F1 Score": [
        f1_score(y_test, dt_pred, average='weighted',
zero_division=0),
        f1_score(y_test, rf_pred, average='weighted',
zero_division=0),
        f1_score(y_test, nb_pred, average='weighted',
zero_division=0)
```

```
    ]
})
print(comparison)
comparison.set_index("Model").plot(kind='bar',
figsize=(10,6), colormap='Set2')
plt.title("Perbandingan Performa Model")
plt.ylabel("Skor")
plt.xticks(rotation=0)
plt.ylim(0, 1)
plt.grid(axis='y')
plt.show()
```

Output yang dihasilkan dari proses evaluasi hasil dan perbandingan sebagaimana gambar berikut :

```
           Model   Akurasi   Presisi    Recall  F1 Score
0  Decision Tree  0.914894  0.926632  0.914894  0.913790
1  Random Forest  0.872340  0.874389  0.872340  0.871756
2    Naive Bayes  0.765957  0.765802  0.765957  0.765531
```



Gambar 7. Output Evaluasi Hasil dan Perbandingan

Tabel 4. Ringkasan Perbandingan

Model	Akurasi	Presisi	Recall	F1 Score
Decision Tree	0.9149	0.9266	0.9149	0.9138
Random Forest	0.8723	0.8743	0.8723	0.8717
Naive Bayes	0.7659	0.7658	0.7659	0.7655

Hasil Klasifikasi

Penelitian terdahulu tentang "Analisis Perbandingan Algoritma *Decision Tree, Random Forest,* dan *Naïve Bayes* untuk Prediksi Banjir di Desa Dayeuhkolot" yang dilakukan oleh Muhammad Bagas Arya Darmawan, Favian Dewanta, dan Sri Astuti, menemukan bahwa algoritma *Random Forest* menunjukkan performa lebih unggul dibandingkan algoritma *Decision Tree* dan *Naïve Bayes* dengan hasil *accuracy* 99,05%,dengan *precision* 97,91%, *recall* 99,18%, dan *f1-score* 98% (Arya Darmawan et al., 2023).

Penelitian serupa juga dilakukan oleh Wida Azis Rahmat, Siti Madinah Ladjamuddin, dan Dipa Teruna Awaludin dengan penelitian berjudul "Perbandingan Algoritma *Decision Tree, Random Forest,* dan *Naïve Bayes* pada Prediksi Penilaian Kepuasan Penumpang Maskapai Pesawat Menggunakan Dataset Kaggle". Hasil analisis menunjukkan bahwa algoritma *Random Forest* berhasil mendapatkan akurasi tertinggi yaitu sebesar 95,44%, sedangkan *Decision Tree* 93,41%, dan *Naïve Bayes* 82,51% (Rahmat et al., 2023).

Performa hasil klasifikasi tidak hanya ditentukan oleh jenis algoritma yang digunakan, namun karakteristik dan jenis data juga mempengaruhi hasil dari klasifikasi. Pada penelitian ini algoritma *Decision Tree* dinilai paling cocok untuk diterapkan pada *Sleep Health and Lifestyle Dataset* karena kemampuannya dalam menganalisis pola dan hubungan antar fitur pada dataset dengan baik, sehingga memberikan hasil klasifikasi lebih optimal dan efektif.

Algoritma *Decision Tree* mempunyai sifat yang transparan dan mudah dipahami. Dengan struktur algoritma berupa pohon keputusan membuatnya menjadi fleksibel dalam mengolah data sehingga mampu menghasilkan performa lebih unggul dan efektif dibandingkan algoritma *Random Forest* dan *Naïve Bayes,* dimana tingkat *accuracy* yang dihasilkan sebesar 91,49%. Selain itu, nilai lainnya seperti *precision, recall,* dan *f1-score* yang dihasilkan juga lebih unggul dan seimbang, menjadikannya mampu untuk menganalisis data mengenai gangguan tidur secara efektif. Sebaliknya, meskipun algoritma *Random Forest* dikenal sebagai algoritma yang stabil dalam menghasilkan prediksi, pada penelitian ini berada pada posisi kedua setelah algoritma *Decision Tree* dengan akurasi sebesar 87,23%. Hal ini bisa jadi disebabkan karena jumlah data yang sedikit dan terbatas, sehingga menghambat proses analisis secara optimal. Sementara algoritma *Naïve Bayes* yang dikenal sebagai algoritma yang cepat dan sederhana, namun pada penelitian ini mendapatkan nilai akurasi terendah yaitu 76,59%. Hal ini terjadi karena algoritma ini menganggap semua variabel yang ada pada dataset tidak saling berhubungan, padahal ada beberapa variabel yang saling terhubung dan saling berinteraksi satu sama lain.

Penutup

Berdasarkan analisis yang dilakukan pada algoritma *Decision Tree, Random Forest,* dan *Naïve Bayes* dalam mengklasifikasikan gangguan tidur, dapat disimpulkan bahwa algoritma *Decision Tree* memiliki kinerja yang optimal dibandingkan algoritma *Random Forest* dan *Naïve Bayes*. Hal ini ditunjukkan oleh hasil *accuracy* tertinggi yang diperoleh pada algoritma *Decision Tree* yaitu sebesar 91.49%, serta kemampuan *precision* dan *recall* yang baik dalam membedakan

penderita gangguan tidur dan tanpa gangguan tidur. Sedangkan algoritma *Random Forest* menghasilkan akurasi sebesar 87.23% dan algoritma *Naïve Bayes* dengan akurasi sebesar 76.59%. Evaluasi performa setiap model dilakukan dengan menggunakan metrik *accuracy, precision, recall,* dan *f1-score* baik secara otomatis melalui *Google Colab* maupun secara manual menggunakan rumus statistik, dan hasilnya menunjukkan konsistensi. Secara keseluruhan, algoritma *Decision Tree* direkomendasikan sebagai algoritma yang efektif dan optimal dalam klasifikasi gangguan tidur berdasarkan data yang digunakan.

Saran

Penelitian ini tentunya masih belum sempurna dan memiliki banyak kekurangan. Diharapkan untuk penelitian – penelitian selanjutnya dapat membandingkan algoritma lainnya untuk mendapatkan hasil analisis yang lebih akurat serta dapat menjadi acuan untuk pengembangan sistem deteksi gangguan tidur yang dapat di implementasikan secara praktis di lingkungan operasional, seperti fasilitas kesehatan atau aplikasi berbasis teknologi.

Ucapan Terimakasih

Ungkapan terimakasih saya haturkan kepada pihak - pihak yang telah membantu dan membimbing tercapainya penelitian ini. Terimakasih kepada seluruh pihak di Fakultas Sains dan Teknologi Universitas Ibrahimy utamanya kepada dosen pembimbing Bapak Lukman Fakih Lidimilah, M.Kom., dan Bapak Ahmad Lutfi, M.Kom,. yang telah membimbing saya sehingga penelitian ini dapat terselesaikan. Serta seluruh teman yang telah mensuport saya.

Daftar Pustaka

Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, *05*, 97–100.

Arya Darmawan, M. B., Dewanta, F., & Astuti, S. (2023). Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot. *TELKA - Telekomunikasi Elektronika Komputasi Dan Kontrol*, *9*(1), 52–61. https://doi.org/10.15575/telka.v9n1.52-61

Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, *5*(2), 640. https://doi.org/10.30865/mib.v5i2.2937

Baharuddin, F., & Tjahyanto, A. (2022). Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Peningkatan Dataset. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *11*(1), 25–31. https://doi.org/10.32736/sisfokom.v11i1.1337

Brianorman, Y., & Sucipto, S. (2024). Prediction of Prospective New Students Using Decision Tree, Random Forest, and Naive Bayes. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(3), 1433–1446. https://doi.org/10.30598/barekengvol18iss3pp1433-1446

Cardova, A., & Hermawan, A. (2023). Implementasi Metode LSTM Untuk Mengklasifikasi Berita Palsu Pada PolitiFact. *Jurnal Fasilkom*, *13*(3), 471–479. https://doi.org/10.37859/jf.v13i3.6175

Dr. H. Fuad Nashori, S.Psi., M.Si., M.Ag., P., & Etik Dwi Wulandari, S.Psi., M.Psi., P. (2017). *Psikologi Tidur: Dari kualitas tidur hingga insomnia*. *11*(1), 1–14.

Fuji Astari, D., Herry Chrisnanto, Y., & Melina, M. (2024). Klasifikasi Tingkat Stres Saat Tidur Menggunakan Algoritma Random Forest. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(5), 3676–3684. https://doi.org/10.36040/jati.v7i5.7750

Hassan, C. A. U., Khan, M. S., & Shah, M. A. (2018). *Comparison of Machine Learning Algorithms in Data classification*. https://doi.org/10.23919/IConAC.2018.8748995

Herikurniawan, B. (n.d.). *Demokrasi Digital: Dinamika Peluang dan Tantangan di Era Media Sosial*. Retrieved 31 October 2024, from https://www.kompasiana.com/bagusheri/65ae322dde948f6e3d4f7272/demokrasi-digital-dinamika-peluang-dan-tantangan-di-era-media-sosial?form=MG0AV3

Kamaliyah, A. F., Studi, P., Informasi, T., Sains, F., Tekologi, D. A. N., Ibrahimy, U., & Situbondo, S. (2023). *Pesantren Dengan Metode Elbow Dan Silhouette Pada Algoritma Clustering K-Medoids Pada Algoritma Clustering K-Medoids*.

Kesehatan, D. J. P. (2022). *Mengenal Gangguan Tidur*.

Mais, F. R., Rompas, S. S. J., & Gannika, L. (2020). Kecanduan Game Online Dengan Insomnia Pada Remaja. *Jurnal Keperawatan*, *8*(2), 18. https://doi.org/10.35790/jkp.v8i2.32318

Prabowo, A., Wardani, S., Wijaya Dewantoro, R., & Wesly, W. (2023). Komparasi Tingkat Akurasi Random Forest dan Decision Tree C4.5 Pada Klasifikasi Data Penyakit Infertilitas. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(1), 218–224. https://doi.org/10.30865/klik.v4i1.1115

Rahmat, A. ., Ladjamuddin, M. ., & Awaludin, T. . (2023). Perbandingan Algoritma Decision Tree, Random Forest Dan Naive Bayes Pada Prediksi Penilaian Kepuasan Penumpang Maskapai Pesawat Menggunakan Dataset Kaggle. *Jurnal Rekayasa Informasi*, *12*(2), 150–159.

Renganathan, V. (2021). *Machine Learning Algorithms for Data Scientists: An Overview*. Vinaitheerthan Renganathan.

Sari, D. (2024). Prediksi Gangguan Tidur pada Sleep Health and Lifestyle Menggunakan Support Vector Machine dan Neural Network. *JAVIT : Jurnal Vokasi Informatika*, 36–42. https://doi.org/10.24036/javit.v4i1.168

Setyawan, D. (2016). Analisis Implementasi Pemanfaatan Sistem Informasi Manajemen Rumah Sakit (Simrs) Pada Rsud Kardinah Tegal. *IJCIT (Indonesian Journal on Computer and …*, *1*(2), 54–61.

Stroke, N. I. of N. D. and. (n.d.). *Brain Basics: Understanding Sleep*.

Yasir, M., & Suraji, R. (2023). Perbandingan Metode Klasifikasi Naive Bayes, Decision, Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 pada Media Sosial Youtube. *Jurnal Cahaya Mandalika (JCM)*, *3*(2), 180–192.

Zed, M. (2008). *Metode Penelitian Kepustakaan (Library Research)*. Yayasan Pustaka Obor Indonesia, 2008.