

# Identifikasi Gangguan Tidur Menggunakan Klasifikasi Berbasis Decision Tree

Fifi Andriani<sup>1</sup>, Nurrahmania<sup>2</sup>, Nur Alfiana<sup>3</sup>.

<sup>1</sup>Ilmu Komputer, Universitas Muhammadiyah Bima, Bima, Indonesia

<sup>2</sup>Ilmu Komputer, Universitas Muhammadiyah Bima, Bima, Indonesia

<sup>3</sup>Ilmu Komputer, Universitas Muhammadiyah Bima, Bima, Indonesia

## Article Info

### Article history:

Received 18-06-2025

Revised 29-06-2025

Accepted 30-06-2025

### Kata Kunci:

Sleep disorder

Sleep apnea

insomnia

klasifikasi

decision tree

## Abstrak (10 PT)

Kondisi gangguan tidur sering kali tidak disadari sejak dini oleh banyak individu, sehingga berpotensi memburuk dan menimbulkan masalah kesehatan fisik maupun psikologis jangka panjang yang serius. Deteksi dini sangat dibutuhkan untuk mengantisipasi dampak negatif lebih lanjut dan mengurangi beban penyakit terkait tidur di masyarakat. Penelitian ini mengembangkan dan mengevaluasi sistem klasifikasi gangguan tidur berbasis algoritma Decision Tree melalui proses pelatihan dan pengujian model, dengan tujuan membantu masyarakat mendeteksi gangguan tidur secara cepat dan efisien. Sistem ini memanfaatkan dataset Sleep Health and Lifestyle dari Kaggle, yang terdiri atas 374 sampel dengan 13 atribut terkait pola tidur dan gaya hidup. Model klasifikasi yang dikembangkan membedakan tiga kategori gangguan: No Disorder, Sleep Apnea, dan Insomnia. Hasil pengujian menunjukkan model mencapai akurasi sebesar 89 persen. Angka akurasi tersebut menunjukkan keandalan metode Decision Tree dalam mengklasifikasikan gangguan tidur sesuai kategorinya. Keunggulan sistem ini terletak pada kemampuannya mengidentifikasi gangguan tidur secara mandiri kapan pun dan di mana pun. Dengan demikian, pengguna tidak perlu terlebih dahulu mengunjungi tenaga kesehatan, sehingga sistem ini diharapkan mempermudah deteksi dini gangguan tidur dan meningkatkan kesadaran masyarakat akan pentingnya menjaga kualitas tidur.

This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.



## Penulis Korespondensi:

Fifi Andriani

Computer Science, Faculty of Engineering and Computer Science

Universitas Muhammadiyah Bima

43400, Sarae, Bima, Indonesia

Email: [fifi@umb.ac.id](mailto:fifi@umb.ac.id)

## 1. PENDAHULUAN (10 PT)

Tidur memiliki peran yang sangat penting dalam menjaga keseimbangan fungsi fisiologis, kognitif, dan emosional seseorang. Kualitas dan kuantitas tidur yang buruk dapat berdampak serius terhadap kesehatan, mulai dari penurunan konsentrasi, gangguan suasana hati, hingga peningkatan risiko penyakit kronis seperti hipertensi, diabetes melitus, dan penyakit jantung koroner (Colten & Altevogt, 2006; Ramar et al., 2021). Selain itu, gangguan tidur juga dapat menurunkan produktivitas, meningkatkan risiko kecelakaan kerja, serta berdampak negatif pada kualitas hidup secara keseluruhan (Karna et al., 2023; Shah et al., 2025). Studi dari American Academy of Sleep Medicine (2021) menyatakan bahwa tidur yang tidak optimal merupakan salah satu faktor risiko utama bagi kesejahteraan fisik dan mental. Namun, yang menjadi perhatian adalah bahwa sebagian besar penderita gangguan tidur, seperti insomnia atau sleep apnea, tidak menyadari bahwa dirinya mengalami gangguan, karena kurangnya informasi dan akses terhadap metode deteksi dini (Colten & Altevogt, 2006; AASM, 2020). Ketidaktahuan masyarakat mengenai indikator gangguan tidur ini berpotensi memperburuk kondisi kesehatan secara akumulatif jika tidak ditangani secara tepat waktu.

Metode diagnosa medis seperti polisomnografi (PSG) memang menjadi standar utama dalam mengidentifikasi gangguan tidur, tetapi prosesnya mahal, invasif, dan tidak praktis untuk penggunaan luas di tingkat masyarakat. Selain itu, keterbatasan fasilitas laboratorium tidur, kebutuhan teknisi ahli, dan waktu observasi yang panjang membuat metode ini kurang efisien sebagai alat skrining awal (Gerstenslager & Slowik, 2023). Sekitar 80–90% penderita sleep apnea, misalnya, tidak terdiagnosis karena hambatan tersebut (IOM, 2006). Oleh karena itu, diperlukan pendekatan berbasis teknologi yang lebih murah, cepat, dan mudah diterapkan untuk membantu masyarakat mengenali gangguan tidur secara mandiri. Perkembangan teknologi pengumpulan data pribadi—seperti pelacak tidur dan perangkat wearable—membuka peluang untuk mengintegrasikan data tersebut ke dalam sistem kecerdasan buatan (AI), termasuk algoritma machine learning, guna membangun model klasifikasi gangguan tidur secara otomatis dan akurat.

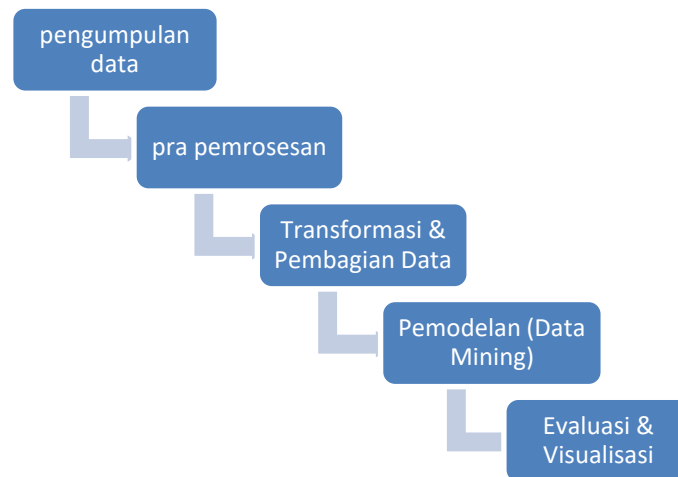
Berbagai studi telah menunjukkan bahwa algoritma machine learning efektif dalam mengklasifikasikan jenis gangguan tidur menggunakan data-data yang dikumpulkan secara pasif. Widasari et al. (2020) menggunakan pendekatan bagged decision tree untuk mengklasifikasikan beberapa gangguan tidur dan memperoleh hasil yang baik. Monowar et al. (2024) juga menunjukkan bahwa pendekatan multi-layered ensemble dapat meningkatkan akurasi klasifikasi menjadi lebih dari 96% pada data sleep disorder. Sementara itu, Zhang et al. (2024) menguji klasifikasi multi-kelas menggunakan data dengan label insomnia, sleep apnea, dan normal, dan menemukan bahwa metode SVM dan regresi logistik dapat mencapai akurasi hingga 91%. Keunggulan algoritma decision tree terletak pada kemampuannya dalam memetakan proses pengambilan keputusan secara intuitif, keterbacaannya oleh manusia, serta ketahanannya terhadap data non-linear dan tidak terdistribusi normal. Oleh karena itu, decision tree menjadi kandidat yang tepat untuk pengembangan sistem klasifikasi gangguan tidur berbasis data.

Penelitian ini menyajikan pendekatan klasifikasi gangguan tidur menggunakan algoritma decision tree berdasarkan studi kasus data sekunder yang diambil dari Kaggle. Dataset tersebut memuat beragam variabel seperti durasi tidur, pekerjaan, jenis kelamin, kualitas tidur, detak jantung, aktivitas harian, berat badan, tingkat stres, tekanan darah, dan jumlah langkah per hari, dengan tiga label utama yaitu sleep apnea, insomnia, dan no disorder. Pemilihan judul *Klasifikasi Gangguan Tidur Menggunakan Algoritma Decision Tree* mencerminkan fokus utama penelitian dalam menyederhanakan proses deteksi gangguan tidur melalui teknik klasifikasi otomatis berbasis data. Dengan menyediakan model skrining yang murah, cepat, dan mudah digunakan, penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan kesadaran masyarakat akan pentingnya kesehatan tidur, serta membantu dalam deteksi dini gangguan yang selama ini sering terabaikan. Pendekatan ini berpotensi menjadi solusi awal yang efektif untuk membantu sistem kesehatan dalam melakukan intervensi lebih dini, sebelum kondisi pasien berkembang menjadi lebih kompleks.

## 2. METODE (10 PT)

Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) sebagai kerangka metode penelitian untuk melakukan proses penemuan pengetahuan dari data gangguan tidur. KDD dipilih karena menyediakan tahapan yang sistematis dalam ekstraksi informasi yang bermakna dari kumpulan data besar, khususnya dalam konteks klasifikasi. Pada tahap data mining, algoritma Decision Tree digunakan untuk membangun model klasifikasi yang mampu mengidentifikasi jenis gangguan tidur berdasarkan atribut-atribut yang tersedia.

Penelitian ini menggunakan kerangka Knowledge Discovery in Databases (KDD) dengan Decision Tree (CART) untuk klasifikasi gangguan tidur. Tahapan utamanya:



**Gambar 1.** Alur Penelitian

### 1. Pengumpulan Data

Data dalam penelitian ini diperoleh dari Sleep Health and Lifestyle Dataset yang terdiri atas 374 rekaman individu. Setiap rekaman memuat informasi yang berkaitan dengan kondisi kesehatan dan gaya hidup responden, meliputi jenis kelamin, usia, serta pekerjaan atau profesi yang dijalani. Selain itu, dataset ini juga mencatat berbagai indikator yang berhubungan dengan kualitas tidur, seperti durasi tidur harian dan penilaian kualitas tidur. Aspek gaya hidup dan kesehatan fisik turut direkam, termasuk tingkat aktivitas fisik, tingkat stres, indeks massa tubuh (BMI), tekanan darah, denyut jantung per menit, serta jumlah langkah yang ditempuh setiap harinya.

Label utama dalam dataset ini adalah sleep disorder, yang menunjukkan ada atau tidaknya gangguan tidur pada masing-masing individu. Pemilihan dataset ini didasarkan pada kelengkapan atribut yang mendukung analisis prediksi gangguan tidur menggunakan metode machine learning, karena mencakup faktor-faktor yang berhubungan langsung dengan kesehatan, aktivitas harian, dan pola tidur responden.

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	120/80	77	4200	No Disorder
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	No Disorder
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	No Disorder
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
5	6	Male	28	Software Engineer	5.9	4	30	8	Obese	140/90	85	3000	Insomnia
6	7	Male	29	Teacher	6.3	6	40	7	Obese	140/90	82	3500	Insomnia
7	8	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	No Disorder
8	9	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	No Disorder
9	10	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	No Disorder

## Gambar 2. Dataset

### 2. Pra pemrosesan

Tahap pra-pemrosesan data dilakukan untuk memastikan dataset yang digunakan bersih dan siap diolah pada proses analisis selanjutnya. Langkah pertama adalah melakukan pembersihan data dengan menghapus seluruh entri yang terdeteksi sebagai duplikat, serta menghilangkan rekaman yang memiliki nilai kosong lebih dari 20% karena dianggap tidak representatif.

Selanjutnya, dilakukan penanganan pada kolom sleep disorder yang semula hanya terdiri atas dua kelas, yaitu insomnia dan sleep apnea. Nilai kosong pada kolom ini diisi dengan kategori baru, yaitu no disorder, sehingga label target pada penelitian ini menjadi tiga kelas. Untuk atribut lain yang memiliki nilai hilang kurang dari 20%, dilakukan proses imputasi agar tidak mengurangi jumlah data yang tersedia. Nilai yang hilang pada atribut numerik diisi menggunakan nilai rata-rata, sedangkan pada atribut kategorik diisi menggunakan nilai modus. Dengan tahapan pra-pemrosesan ini, dataset menjadi lebih bersih dan lengkap sehingga dapat digunakan secara optimal dalam proses analisis dan pemodelan.

### 3. Transformasi & Pembagian Data

Setelah melalui tahap pembersihan dan imputasi, data selanjutnya diproses melalui tahap transformasi agar siap digunakan pada pemodelan. Seluruh atribut numerik dinormalisasi menggunakan teknik min-max scaling untuk menyamakan rentang nilai setiap fitur sehingga model machine learning dapat memproses data secara optimal. Pada tahap ini, rekayasa fitur (feature engineering) juga dapat dilakukan secara opsional untuk mengekstraksi informasi tambahan yang berpotensi meningkatkan kinerja model, misalnya dengan membentuk rasio tidur gelombang lambat (slow wave sleep ratio) dari data yang tersedia.

Setelah proses transformasi selesai, dataset kemudian dibagi menjadi dua bagian untuk keperluan pelatihan dan pengujian model. Sebanyak 80% data digunakan sebagai data latih, sedangkan 20% sisanya digunakan sebagai data uji. Pembagian data dilakukan secara stratified agar distribusi kelas pada label sleep disorder tetap seimbang di kedua subset, sehingga hasil evaluasi model lebih representatif.

### 4. Pemodelan (Data Mining)

Tahap pemodelan dalam penelitian ini menggunakan algoritma Decision Tree dengan pendekatan Classification and Regression Tree (CART). Untuk memperoleh kinerja model yang optimal, dilakukan proses penyetelan parameter (hyperparameter tuning) menggunakan grid search yang dipadukan dengan 5-fold cross validation. Parameter utama yang disesuaikan meliputi kedalaman maksimum pohon keputusan (max\_depth) dan jumlah minimum sampel pada daun (min\_samples\_leaf).

Setelah diperoleh kombinasi parameter terbaik, model dilatih menggunakan data latih yang telah disiapkan sebelumnya, kemudian dievaluasi kinerjanya pada data uji. Tahap ini memastikan model mampu melakukan prediksi terhadap kelas sleep disorder secara akurat dan dapat digunakan sebagai dasar analisis lebih lanjut.

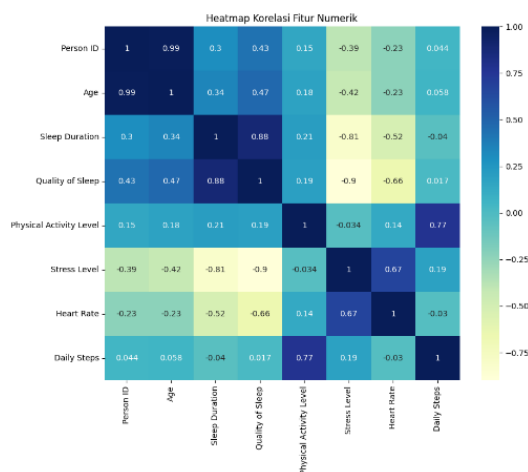
### 5. Evaluasi & Visualisasi

Kinerja model dievaluasi menggunakan beberapa metrik klasifikasi, yaitu akurasi, precision, recall, dan F1-score. Selain itu, confusion matrix juga digunakan untuk menilai seberapa baik model mengklasifikasikan setiap kelas sleep disorder secara rinci. Evaluasi ini memberikan gambaran menyeluruh mengenai kemampuan model dalam membedakan kelas insomnia, sleep apnea, dan no disorder.

Tahap interpretasi dilakukan dengan menganalisis struktur cabang pohon keputusan yang dihasilkan. Analisis ini bertujuan untuk mengidentifikasi fitur-fitur yang paling berpengaruh dalam menentukan prediksi gangguan tidur, sehingga hasil penelitian tidak hanya memberikan model prediksi, tetapi juga wawasan tentang faktor-faktor utama yang memengaruhi kualitas tidur. Sebagai pendukung visualisasi, alur proses Knowledge Discovery in Database (KDD) ditampilkan untuk

menggambarkan tahapan penelitian, mulai dari pengumpulan data hingga evaluasi model. Selain itu, visualisasi pohon keputusan juga disajikan agar pembaca dapat memahami proses pengambilan keputusan model secara lebih jelas dan interpretatif.

### 3. HASIL DAN PEMBAHASAN (10 PT)



**Gambar 3.** Heatmap Korelas

Gambar yang Anda tampilkan merupakan heatmap korelasi antar fitur numerik pada dataset Sleep Health and Lifestyle. Heatmap ini menunjukkan kekuatan dan arah hubungan linear antarvariabel numerik yang ada pada dataset. Warna biru tua menunjukkan korelasi positif yang tinggi (mendekati +1), sedangkan warna kuning muda hingga hijau menunjukkan korelasi negatif (mendekati -1). Nilai 0 menunjukkan hampir tidak ada korelasi.

#### Person ID dan Age

Memiliki korelasi sangat tinggi (0,99) karena kemungkinan Person ID diurutkan berdasarkan usia sehingga hampir identik. Variabel Person ID sebaiknya tidak digunakan sebagai fitur karena tidak informatif untuk prediksi.

#### Sleep Duration dan Quality of Sleep

Memiliki korelasi positif yang kuat (0,88), artinya semakin lama seseorang tidur, cenderung kualitas tidurnya semakin baik.

#### Stress Level dan Sleep Duration

Terdapat korelasi negatif yang kuat (-0,81), menunjukkan bahwa semakin tinggi tingkat stres, semakin pendek durasi tidur seseorang.

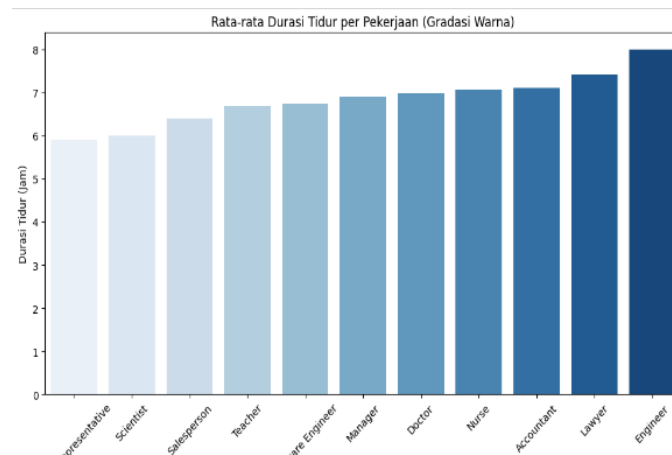
#### Daily Steps dan Physical Activity Level

Korelasi positif tinggi (0,77), artinya semakin banyak langkah harian, semakin tinggi pula tingkat aktivitas fisik yang dicatat.

#### Heart Rate dan Quality of Sleep

Korelasi negatif sedang (sekitar -0,66), mengindikasikan bahwa denyut jantung yang lebih tinggi cenderung berhubungan dengan kualitas tidur yang lebih rendah.

Dari heatmap ini, dapat disimpulkan fitur-fitur yang memiliki korelasi sangat tinggi mungkin redundan (multicollinearity) dan perlu dipertimbangkan saat pemodelan, misalnya dengan menghapus fitur Person ID atau memilih salah satu di antara fitur yang terlalu berkorelasi tinggi.



**Gambar 4.** Sleep Duration

Menggambarkan perbandingan rata-rata durasi tidur responden berdasarkan jenis pekerjaan yang dimiliki. Visualisasi ini memperlihatkan adanya variasi yang cukup mencolok antar profesi terkait jumlah waktu tidur yang diperoleh tiap malam. Kelompok pekerjaan dengan durasi tidur terpendek ditemukan pada profesi Sales Representative, dengan rata-rata 5,9 jam per malam. Disusul oleh profesi Scientist (6,0 jam), Salesperson (6,4 jam), dan Teacher (6,7 jam) yang juga menunjukkan durasi tidur relatif lebih singkat.

Sementara itu, peningkatan rata-rata waktu tidur mulai tampak pada profesi seperti Software Engineer (6,8 jam), Manager (6,9 jam), dan Doctor (7,0 jam). Durasi tidur yang lebih tinggi tercatat pada profesi Nurse dan Accountant, masing-masing dengan rata-rata 7,1 jam, serta Lawyer dengan 7,4 jam. Puncaknya, profesi Engineer menunjukkan waktu tidur terpanjang, yaitu 8,0 jam per malam. Perbedaan ini mengisyaratkan kemungkinan adanya keterkaitan antara jenis pekerjaan dan kebiasaan tidur. Pekerjaan dengan beban kerja tinggi atau jadwal yang tidak teratur, seperti profesi di bidang penjualan dan pendidikan, cenderung dikaitkan dengan waktu tidur yang lebih singkat. Sebaliknya, profesi yang memungkinkan pengelolaan waktu kerja yang lebih konsisten, seperti Engineer dan Lawyer, menunjukkan kecenderungan memiliki durasi tidur yang lebih panjang. Hal ini menunjukkan bahwa faktor pekerjaan berperan penting dalam memengaruhi kualitas dan kuantitas tidur, sehingga dapat menjadi variabel yang perlu diperhatikan dalam studi terkait kesehatan tidur pada gambar 4.

```

... Jumlah total data : 374
   Jumlah data training : 299
   Jumlah data testing : 75

=== Data Training (X_train) ===
  Person ID  Gender  Age  Occupation  Sleep Duration  Quality of Sleep \
192         193     1   43           7             6.5             6
75          76     1   33           1             6.0             6
84          85     1   35           9             7.5             8
362        363     0   59           5             8.2             9
16         17     0   29           5             6.5             5
..         ...     ... ..           ...           ...             ...
71          72     1   33           1             6.1             6
106        107     0   37           5             6.1             6
270        271     0   49           5             6.1             6
348        349     0   57           5             8.2             9
102        103     0   36           10            7.2             8

  Physical Activity Level  Stress Level  BMI Category  Blood Pressure \
192                    45             7             3             15
75                     30             8             0             9
84                     60             5             1             6
362                    75             3             3             23
16                     40             7             1             18
..                     ...             ...           ...           ...

```

**Gambar 5.** Split Dataset

Pembagian data dilakukan dengan tujuan memisahkan dataset menjadi dua bagian, yaitu data pelatihan (training set) dan data pengujian (testing set). Dari total 374 data yang tersedia, sebanyak 299 data (sekitar 80%) dialokasikan untuk proses pelatihan model, sedangkan 75 data sisanya (sekitar 20%) digunakan untuk menguji kinerja model yang telah dibangun. Pembagian ini dirancang untuk memberikan kesempatan bagi model dalam mempelajari pola dari sebagian besar data, sekaligus menguji kemampuannya dalam melakukan prediksi terhadap data yang belum pernah digunakan sebelumnya, guna menilai kualitas generalisasi model.

Data pelatihan ( $X_{train}$ ) terdiri atas sejumlah fitur yang menggambarkan karakteristik masing-masing individu dalam dataset. Fitur-fitur tersebut antara lain:

- 1) Gender dan Age, yang merepresentasikan data demografis;
- 2) Occupation, menunjukkan jenis pekerjaan yang berpotensi memengaruhi pola tidur dan tingkat stres;
- 3) Sleep Duration serta Quality of Sleep, sebagai indikator utama terkait kualitas tidur;
- 4) Physical Activity Level dan Stress Level, yang mencerminkan gaya hidup dan keseimbangan aktivitas harian responden;
- 5) BMI Category dan Blood Pressure, sebagai indikator kesehatan yang relevan terhadap kemungkinan gangguan tidur

```

... === Classification Report ===
              precision    recall  f1-score   support

   Insomnia         0.76      0.81      0.79         16
  No Disorder         0.93      1.00      0.97         43
   Sleep Apnea         0.92      0.69      0.79         16

 accuracy                   0.89         75
 macro avg              0.87      0.83      0.85         75
 weighted avg           0.89      0.89      0.89         75

```

**Gambar 6.** Evaluasi Model

Evaluasi kinerja model klasifikasi dilakukan untuk mengukur sejauh mana model dapat mengenali dan membedakan jenis gangguan tidur berdasarkan fitur-fitur input. Pengujian dilakukan menggunakan testing set yang terdiri dari 75 sampel (20% dari total data), yang sebelumnya belum pernah digunakan dalam proses pelatihan. Evaluasi dilakukan melalui dua pendekatan utama, yaitu classification report dan confusion matrix. Classification report memberikan metrik evaluasi berupa precision, recall, dan f1-score untuk masing-masing kelas target: Insomnia, No Disorder, dan Sleep Apnea.

Hasil evaluasi menunjukkan bahwa model memiliki akurasi keseluruhan sebesar 89%, dengan performa klasifikasi terbaik pada kelas No Disorder, yang memperoleh precision dan recall sebesar 0.93 dan 1.00 secara berurutan. Hal ini menunjukkan bahwa model sangat andal dalam mengidentifikasi individu yang tidak mengalami gangguan tidur. Untuk kelas Insomnia, precision sebesar 0.76 dan recall sebesar 0.81 menunjukkan bahwa model cukup mampu mengenali individu yang mengalami gangguan tersebut, meskipun terdapat beberapa kesalahan klasifikasi. Sementara itu, kelas Sleep Apnea memperoleh precision sebesar 0.92, tetapi recall hanya 0.69, menandakan bahwa sebagian kasus sleep apnea masih belum dapat dikenali dengan optimal oleh model.

Secara umum, nilai macro average dari precision, recall, dan f1-score masing-masing sebesar 0.87, 0.83, dan 0.85, sementara weighted average dari ketiga metrik tersebut adalah 0.89. Hal ini

mengindikasikan bahwa model memiliki performa yang cukup seimbang di seluruh kelas, meskipun masih terdapat ruang perbaikan khususnya pada kelas minoritas seperti Sleep Apnea. Hasil ini diperkuat oleh visualisasi confusion matrix, yang menggambarkan distribusi prediksi model terhadap label sebenarnya. Dari 16 data aktual Insomnia, sebanyak 13 diklasifikasikan dengan benar, sementara 3 sisanya salah diklasifikasikan sebagai kelas lain. Untuk Sleep Apnea, hanya 11 dari 16 data yang diprediksi dengan benar, sedangkan sisanya diklasifikasikan sebagai Insomnia atau No Disorder. Sebaliknya, seluruh 43 data No Disorder berhasil dikenali secara akurat oleh model tanpa adanya kesalahan klasifikasi.

Ketidakeimbangan dalam performa antar kelas ini kemungkinan disebabkan oleh distribusi data yang tidak merata (class imbalance), serta kemungkinan adanya kemiripan pola data antara kelas Insomnia dan Sleep Apnea. Untuk mengatasi hal ini, langkah-langkah lanjutan seperti penyeimbangan dataset (oversampling pada kelas minoritas), pemilihan algoritma klasifikasi yang lebih kompleks, atau teknik feature engineering tambahan dapat dipertimbangkan untuk meningkatkan kemampuan generalisasi model terhadap seluruh kelas target.

#### 4. KESIMPULAN (10 PT)

Penelitian ini menggunakan metode klasifikasi yang bertujuan untuk memprediksi masalah gangguan tidur berdasarkan tiga label yang telah ditentukan, data Pemodelan yang telah dibangun sudah bisa mengidentifikasi gangguan tidur dengan memanfaatkan algoritma klasifikasi Decision Tree menggunakan data gaya hidup individu yang didapatkan dari kaggle. Tiga jenis kondisi tidur yang diidentifikasi adalah Insomnia, Sleep Apnea, dan No Disorder. Berdasarkan hasil pelatihan dan pengujian model, algoritma Decision Tree mampu mencapai tingkat akurasi sebesar 89%, yang menunjukkan performa yang sangat baik dalam mengklasifikasikan jenis gangguan tidur. Analisis fitur menunjukkan bahwa variabel durasi tidur, tingkat stres, dan aktivitas fisik memiliki pengaruh signifikan terhadap kemungkinan seseorang mengalami gangguan tidur. Individu dengan durasi tidur yang tidak mencukupi, tingkat stres yang tinggi, serta aktivitas fisik yang rendah cenderung lebih rentan mengalami gangguan seperti insomnia dan sleep apnea. Model klasifikasi berbasis Decision Tree tidak hanya memberikan hasil klasifikasi yang cukup akurat, tetapi juga menawarkan interpretasi yang mudah dipahami, sehingga dapat menjadi alat bantu yang potensial dalam proses deteksi awal gangguan tidur. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam upaya peningkatan kesadaran dan pencegahan dini terhadap gangguan tidur melalui pendekatan data-driven yang efisien dan cepat sehingga dapat memberikan pencegahan awal yang cepat.

#### DAFTAR PUSTAKA (10 PT)

- [1] M. A. Alasmary, M. Z. A. Bhuiyan, and G. Wang, "Sleep disorder detection using machine learning techniques: A review," *IEEE Access*, vol. 9, pp. 14682–14704, 2021.
- [2] J. W. Kim et al., "Classification of sleep disorders based on heart rate variability using deep neural networks," *Sensors*, vol. 20, no. 4, pp. 1–15, 2020.
- [3] R. T. Nguyen et al., "Predicting sleep quality using wearable sensors and machine learning," *J. Biomed. Inform.*, vol. 105, p. 103411, 2020.
- [4] Y. Yu, M. Li, L. Liu, Y. Li, and J. Wang, "Clinical big data and deep learning: Applications, challenges, and future outlooks," *Big Data Min. Anal.*, vol. 2, no. 4, pp. 288–305, 2019, doi: 10.26599/BDMA.2019.902000
- [5] G. Nguyen et al., "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.
- [6] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
- [7] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.
- [8] F. Al-Turjman, H. Zahmatkesh, and L. Mostarda, "Quantifying uncertainty in internet of medical things and big-data services using intelligence and deep learning," *IEEE Access*, vol. 7, pp. 115749–115759, 2019, doi: 10.1109/ACCESS.2019.2931637.

- [9] S. Kumar and M. Singh, "Big data analytics for healthcare industry: Impact, applications, and tools," *Big Data Min. Anal.*, vol. 2, no. 1, pp. 48–57, 2019, doi: 10.26599/BDMA.2018.9020031.
- [10] D. Nallaperuma et al., "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4679–4690, 2019, doi: 10.1109/TITS.2019.2924883.
- [11] J. R. Saura, B. R. Herraiez, and A. Reyes-Menendez, "Comparing a traditional approach for financial brand communication analysis with a big data analytics technique," *IEEE Access*, vol. 7, pp. 37100–37108, 2019, doi: 10.1109/ACCESS.2019.2905301.
- [12] K. Sivaraman, R. M. V. Krishnan, B. Sundarraj, and S. S. Gowthem, "Network failure detection and diagnosis by analyzing syslog and SNS data: Applying big data analysis to network operations," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 9S3, pp. 883–887, 2019, doi: 10.35940/ijitee.I3187.0789S319.
- [13] A. D. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors*, vol. 19, no. 2, pp. 1–17, 2019, doi: 10.3390/s19020326.
- [14] S. Leonelli and N. Tempini, *Data Journeys in the Sciences*. Cham, Switzerland: Springer, 2020.
- [15] C. Shang and F. You, "Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era," *Engineering*, vol. 5, no. 6, pp. 1010–1016, 2019, doi: 10.1016/j.eng.2019.01.019.