

## Studi Perbandingan Algoritma Random Forest dan K-Nearest Neighbors (KNN) dalam Klasifikasi Gangguan Tidur

Nurul Khasanah<sup>1\*</sup>, Daniati Uki Eka Saputri<sup>2</sup>, Faruq Aziz<sup>3</sup>, Taopik Hidayat<sup>4</sup>

<sup>1,2,3,4</sup>Universitas Nusa Mandiri

Jl. Raya Jatiwaringin No.2, RT.8/RW.13, Cipinang Melayu, Kec. Makasar,  
Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13620

e-mail: <sup>1</sup>nurul.nuk@nusamandiri.ac.id, <sup>2</sup>daniati.due@nusamandiri.ac.id, <sup>3</sup>faruq.fqs@nusamandiri.ac.id,  
<sup>4</sup>taopik.toi@nusamandiri.ac.id

(\* ) Corresponding Author

Artikel Info : Diterima : 10-09-2024 | Direvisi : 06-12-2024 | Disetujui : 21-01-2025

**Abstrak** - Gangguan tidur seperti insomnia dan sleep apnea, dapat memengaruhi kualitas hidup secara signifikan serta meningkatkan risiko penyakit kronis, termasuk hipertensi dan diabetes. Identifikasi dini dan klasifikasi gangguan tidur menjadi langkah penting dalam mencegah dampak kesehatan yang lebih lanjut. Penelitian ini bertujuan untuk membandingkan performa algoritma Random Forest dan K-Nearest Neighbors (KNN) dalam klasifikasi gangguan tidur menggunakan *Sleep Health and Lifestyle Dataset* dari Kaggle. Dataset ini berisi data kesehatan dan gaya hidup yang relevan dengan pola tidur dan kategori gangguan tidur, yaitu 'None', 'Sleep Apnea', dan 'Insomnia'. Random Forest menghasilkan akurasi sebesar 89,69%, dengan performa terbaik pada kategori 'None', yang mencapai recall 96,08%. Sebaliknya, KNN menunjukkan akurasi sebesar 87,02% dengan parameter  $K=5$ . Hasil penelitian ini menunjukkan bahwa algoritma Random Forest lebih unggul dibandingkan KNN, khususnya dalam klasifikasi kategori 'None'. Namun, tantangan signifikan ditemukan dalam mendeteksi kategori 'Sleep Apnea', dengan recall hanya sebesar 74,55% pada Random Forest. Hal ini diduga disebabkan oleh ketidakseimbangan data, yang memengaruhi kemampuan algoritma untuk mengenali kategori dengan data minoritas. Penelitian ini menegaskan bahwa Random Forest lebih efektif dalam klasifikasi gangguan tidur dibandingkan KNN. Untuk meningkatkan performa deteksi, langkah penelitian selanjutnya mencakup penyeimbangan data menggunakan teknik seperti *oversampling* atau *undersampling* serta eksplorasi algoritma lain, seperti XGBoost atau ensemble methods lainnya. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknologi untuk diagnosis gangguan tidur secara lebih akurat.

Kata Kunci : Klasifikasi, Gangguan Tidur, Random Forest, KNN

**Abstracts** - Sleep disorders, such as insomnia and sleep apnea, can significantly affect quality of life and increase the risk of chronic diseases, including hypertension and diabetes. Early identification and classification of sleep disorders are crucial steps in preventing further health impacts. This study aims to compare the performance of the Random Forest and K-Nearest Neighbors (KNN) algorithms in classifying sleep disorders using the *Sleep Health and Lifestyle Dataset* from Kaggle. This dataset contains health and lifestyle data relevant to sleep patterns and sleep disorder categories: 'None', 'Sleep Apnea,' and 'Insomnia.' Random Forest achieved an accuracy of 89.69%, with its best performance in the 'None' category, attaining a recall of 96.08%. In contrast, KNN achieved an accuracy of 87.02% with a parameter of  $K=5$ . The results of this study indicate that the Random Forest algorithm outperforms KNN, particularly in classifying the 'None' category. However, significant challenges were found in detecting the 'Sleep Apnea' category, with a recall of only 74.55% in Random Forest. This is suspected to be due to data imbalance, which affects the algorithm's ability to recognize minority class data. This study confirms that Random Forest is more effective in classifying sleep disorders compared to KNN. To improve detection performance, future research steps include balancing the data using techniques such as *oversampling* or *undersampling* and exploring other algorithms, such as XGBoost or ensemble methods. Therefore, this research is expected to contribute to the development of more accurate technology for diagnosing sleep disorders.

Keywords : Classification, Sleep Disorders, Random Forest, KNN



**PENDAHULUAN**

Tidur adalah salah satu aspek vital dari kesehatan manusia yang mempengaruhi kualitas hidup secara keseluruhan (Matur et al., 2021). Gangguan tidur, seperti insomnia, *sleep apnea*, dan sindrom kaki gelisah, tidak hanya berdampak pada kualitas tidur tetapi juga pada kesehatan fisik dan mental seseorang. Gangguan tidur dapat meningkatkan risiko penyakit kronis seperti hipertensi, diabetes, dan depresi, serta mempengaruhi performa sehari-hari dan kualitas hidup secara keseluruhan (Rasmianti et al., 2023).

Dalam era digital yang terus berkembang, pengumpulan data kesehatan dan gaya hidup kini menjadi semakin mudah, luas, dan terintegrasi dengan berbagai teknologi. Data tersebut sering kali mencakup informasi yang beragam, seperti pola tidur, kebiasaan makan, tingkat aktivitas fisik, penggunaan perangkat elektronik, hingga faktor-faktor psikologis dan lingkungan yang turut memengaruhi kualitas tidur. Melalui pengumpulan data yang kaya dan mendalam ini, terdapat potensi besar untuk memanfaatkan teknik analisis data canggih, termasuk pembelajaran mesin dan kecerdasan buatan, guna mengidentifikasi serta mengklasifikasikan gangguan tidur dengan tingkat akurasi yang lebih tinggi. Penerapan teknologi ini tidak hanya mampu mendeteksi pola-pola tersembunyi dalam data, tetapi juga memberikan prediksi yang dapat mendukung langkah intervensi medis yang lebih cepat, tepat, dan berbasis bukti.

Beberapa tahun terakhir, teknologi pembelajaran mesin (*machine learning*) telah banyak digunakan dalam bidang kesehatan untuk menganalisis dan memprediksi kondisi kesehatan individu berdasarkan data yang dikumpulkan. Algoritma klasifikasi seperti Random Forest dan K-Nearest Neighbors (KNN) sering digunakan untuk menganalisis data medis karena kemampuan mereka dalam menangani data yang kompleks serta memberikan prediksi yang akurat (Prasetyo, 2024).

Berbagai penelitian telah dilakukan untuk mengidentifikasi gangguan tidur menggunakan metode analisis data dan pembelajaran mesin. Penelitian oleh (Alvi Norma Utami, 2020) mengenai klasifikasi gangguan tidur *REM behaviour disorder* berdasarkan sinyal EEG menggunakan *machine learning*. Hasil klasifikasi gangguan tidur RBD berdasarkan sinyal EEG menggunakan algoritma ELM menunjukkan bahwa algoritma ELM bisa mengklasifikasikan gangguan tidur RBD dengan non RBD berdasarkan sinyal EEG dengan nilai rata-rata akurasi sebesar mencapai  $70.71\% \pm 5.44$ . Lebih lanjut, studi oleh (Prabowo et al., 2020) mengenai penggunaan algoritma naïve bayes untuk Klasifikasi Tingkat Gangguan Tidur. Hasil validasi metode Naïve Bayes menunjukkan bahwa 80% data akurat dan akan dibandingkan antara 10 data uji diagnostik dan 30 data pengujian. Penelitian lain dilakukan oleh (Widasari et al., 2020) dengan menggunakan Decision-Tree-Based Support Vector Machine (DTB-SVM) untuk klasifikasi gangguan tidur. Berdasarkan hasil eksperimen, metode yang diusulkan memiliki sensitivitas 84,01%, spesifisitas 94,17%, akurasi keseluruhan 86,27%, dan kappa Cohen 0,70. Dalam jurnal Prediksi Kualitas Tidur: Pendekatan *Machine Learning* yang Mengintegrasikan Faktor Kesehatan dan Lingkungan (Putra & Hidayat, 2024) dihasilkan penelitian menggunakan data *National Poll on Healthy Aging (NPHA)*. Nilai akurasi yang dihasilkan yaitu 94,00% dan 44,44% akurasi yang didapatkan dengan menggunakan data latih.

Penelitian ini bertujuan untuk membandingkan performa kedua algoritma tersebut dalam klasifikasi gangguan tidur menggunakan dataset dari Kaggle. Random Forest dikenal karena kemampuannya dalam menangani data yang memiliki banyak fitur serta ketahanannya terhadap *overfitting*. Sementara itu, KNN merupakan algoritma sederhana namun efektif yang bekerja berdasarkan kedekatan data dengan tetangganya (Sukmo et al., 2024). Keduanya diimplementasikan menggunakan tools Rapid Miner, yang memudahkan proses analisis data dengan antarmuka yang intuitif.

Dengan membandingkan algoritma Random Forest dan K-Nearest Neighbors (KNN), penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai algoritma yang lebih unggul dalam melakukan klasifikasi gangguan tidur berdasarkan data kesehatan dan gaya hidup individu. Wawasan ini tidak hanya memberikan kontribusi pada pemahaman teknis mengenai performa masing-masing algoritma, tetapi juga membuka peluang untuk mengembangkan sistem diagnosis otomatis yang lebih andal dan akurat. Sistem tersebut diharapkan mampu mendukung proses deteksi dini gangguan tidur, seperti insomnia dan *sleep apnea*, yang pada akhirnya memungkinkan pemberian intervensi medis yang lebih tepat, efisien, dan cepat. Dengan demikian, penelitian ini tidak hanya berkontribusi pada ranah akademik dan teknologi, tetapi juga memiliki potensi dampak langsung yang signifikan dalam meningkatkan kualitas hidup individu serta mengurangi risiko komplikasi kesehatan yang lebih serius akibat gangguan tidur yang tidak terdiagnosis.

**METODE PENELITIAN**

Untuk metode yang digunakan dalam penelitian ini adalah sebagai berikut:



Sumber: Hasil Penelitian (2024)

Gambar 1. Metode Penelitian

### 1. Akuisisi Dataset

Akuisisi dataset adalah proses sistematis untuk mengumpulkan informasi dari berbagai sumber atau variabel dengan tujuan memperoleh data yang diperlukan untuk analisis, penelitian, atau pembuatan keputusan. Proses ini melibatkan pengumpulan data melalui berbagai metode seperti survei, wawancara, pengamatan, atau penggunaan alat dan teknologi untuk memastikan informasi yang diperoleh relevan, akurat, dan bermanfaat (Cohen et al., 2020).

Penelitian ini menggunakan data set *Sleep Health and Lifestyle Dataset* yang diunduh dari platform Kaggle (Sleep Health and Lifestyle Dataset, n.d.) dengan link <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset?resource=download>. Data ini terdiri dari informasi kesehatan dan gaya hidup yang berkaitan dengan pola tidur. Beberapa variabel penting dalam data ini meliputi *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*. Data set ini terdiri dari 374 data observasi dengan 15 atribut yang relevan.

### 2. Preprocessing Data

Data *preprocessing* adalah tahap awal dalam pengolahan data yang melibatkan serangkaian langkah untuk membersihkan, mengorganisir, dan mempersiapkan data agar siap untuk analisis lebih lanjut (Tang et al., 2020). Proses ini melibatkan berbagai langkah yang penting untuk memastikan data yang digunakan dalam model analisis berkualitas tinggi. Dengan preprocessing yang tepat, data akan lebih mudah dianalisis dan menghasilkan hasil yang lebih akurat. Sebelum menerapkan algoritma Random Forest dan K-Nearest Neighbors (KNN), dilakukan beberapa langkah preprocessing untuk memastikan kualitas data yang digunakan dalam klasifikasi. Langkah-langkah tersebut meliputi:

- a. Penanganan *Missing Values*: Observasi yang memiliki *missing values* ditangani dengan menggunakan metode imputasi atau dihapus jika jumlahnya signifikan.
- b. Normalisasi dan Standardisasi: Atribut numerik seperti durasi tidur dinormalisasi untuk memastikan setiap fitur berada pada skala yang sama, sehingga mencegah algoritma bias terhadap fitur tertentu.
- c. Pengkodean Variabel Kategorikal: Atribut kategorikal seperti BMI Category dikonversi menjadi bentuk numerik menggunakan teknik One-Hot Encoding.
- d. Penghapusan atribut yang tidak berpengaruh dalam proses klasifikasi, seperti jenis kelamin.

### 3. Pembagian Data

Data set dibagi menjadi dua subset: 70% (262 data) digunakan untuk pelatihan model (*training set*) dan 30% (112 data) sisanya untuk pengujian model (*test set*). Pembagian antara data pelatihan dan data pengujian merupakan faktor penting dalam menentukan akurasi model. Kesalahan dalam menentukan proporsi kedua jenis data tersebut dapat memengaruhi nilai akurasi dan presisi yang dihasilkan (Azmi et al., 2023). Pembagian ini dilakukan secara acak untuk memastikan bahwa model yang dihasilkan tidak overfitting dan mampu melakukan generalisasi dengan baik pada data baru.

4. Penerapan Algoritma Random Forest dan KNN

a. Random Forest

Random Forest adalah teknik machine learning yang menggabungkan sejumlah Decision Tree untuk tujuan klasifikasi, di mana keputusan akhir diambil berdasarkan suara mayoritas dari semua pohon keputusan tersebut (Larasati et al., 2022). Setelah data di-*preprocessing*, algoritma Random Forest diterapkan untuk mengklasifikasikan gangguan tidur berdasarkan fitur-fitur yang ada. Random Forest dipilih karena kemampuannya dalam menangani data dengan banyak variabel dan mengurangi risiko *overfitting* melalui penggunaan metode ensemble. Konfigurasi parameter utama dari Random Forest yang digunakan dalam penelitian ini adalah sebagai berikut:

- 1) Jumlah Pohon (n\_estimators): 10 pohon keputusan.
- 2) Kedalaman Maksimal Pohon (max\_depth): Tidak dibatasi, memungkinkan pohon tumbuh hingga setiap daun mengandung data dengan kelas yang homogen.
- 3) Kriteria Pemisahan (criterion): Gini index, untuk mengukur impurity pada setiap node.

b. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah salah satu algoritma *supervised learning* yang digunakan untuk klasifikasi dan regresi (Maulida & Nasution, 2024). Parameter penting dalam KNN adalah nilai K, yaitu jumlah tetangga terdekat yang akan diperiksa oleh algoritma untuk menentukan kelas suatu data. Pemilihan nilai K yang tepat sangat mempengaruhi kinerja model, di mana nilai K yang terlalu kecil dapat menyebabkan model menjadi sensitif terhadap noise, sementara nilai K yang terlalu besar dapat membuat model kurang responsif terhadap pola lokal dalam data. Oleh karena itu, eksperimen dan evaluasi terhadap berbagai nilai K sering dilakukan untuk menemukan nilai yang optimal guna menghasilkan akurasi yang terbaik.

Konfigurasi parameter utama dari K-Nearest Neighbors (KNN) yang digunakan dalam penelitian ini adalah

- 1) K = 5

5. Evaluasi Model

Kinerja model dievaluasi menggunakan *confusion matrix*. *Confusion matrix* adalah sebuah alat evaluasi yang digunakan dalam pembelajaran mesin dan statistik untuk menilai kinerja model klasifikasi (Krstinić et al., 2020). Matriks ini menunjukkan perbandingan antara prediksi model dan nilai sebenarnya dari data. *Confusion matrix* membantu dalam mengevaluasi berbagai metrik kinerja model, seperti akurasi, presisi, recall, dan F1 score, serta memberikan wawasan tentang kekuatan dan kelemahan model dalam klasifikasi (Madani et al., 2024).

**HASIL DAN PEMBAHASAN**

1. Data Set

Berikut adalah sample untuk dataset yang digunakan

Tabel 1. Data set

Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level
1	Male	27	Software Engineer	6.1	6	42
Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder	
6	Overweight	126/83	77	4200	None	

Sumber: (Sleep Health and Lifestyle Dataset, n.d.)(Prasetyo, 2024)

2. Preprocessing Data

Berikut adalah tampilan dataset setelah dilakukan preprocessing

Tabel 2. Data set hasil preprocessing

Person ID	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level
1	27	Software Engineer	6.1	6	42
Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
6	3	126/83	77	4200	None

Sumber: Hasil Penelitian (2024)

3. Pembagian Data

Pembagian dataset dilakukan untuk memastikan bahwa model yang dibangun dapat dilatih dan diuji secara efektif, sehingga hasil yang diperoleh dapat diandalkan dan generalizable. Berikut adalah tabel pembagian dataset:

Tabel 3. Pembagian data set

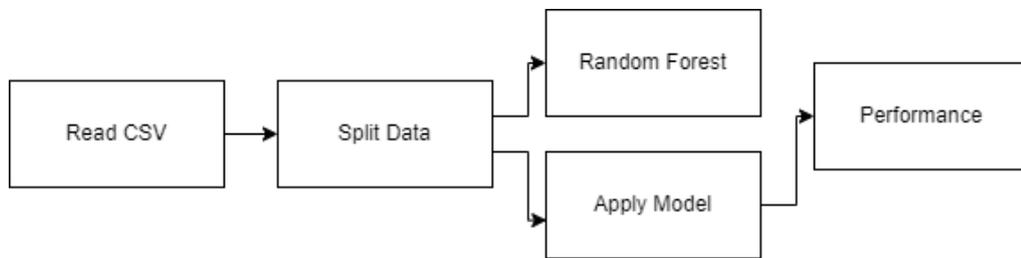
Jumlah Data	:	374
Data Training	:	262
Data Testing	:	112

Sumber: Hasil Penelitian (2024)

4. Penerapan Algoritma Random Forest dan K-Nearest Neighbors (KNN)

a. Random Forest

Dalam proses penerapan algoritma Random Forest dalam klasifikasi gangguan tidur menggunakan aplikasi RapidMiner. Berikut adalah proses yang dilakukan menggunakan aplikasi tersebut:



Sumber: Hasil Penelitian (2024)

Gambar 2. Penerapan algoritma Random Forest menggunakan Rapid Miner

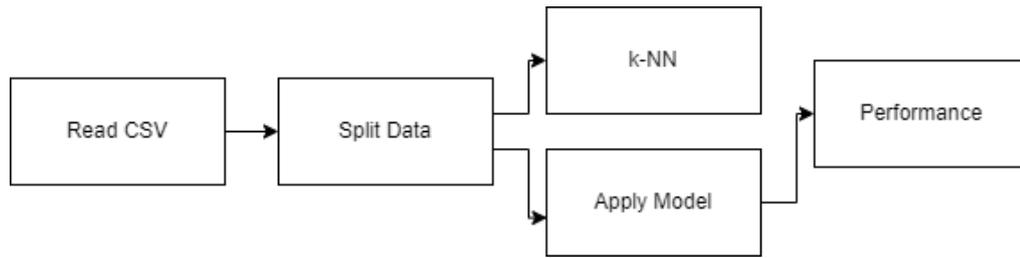
Parameter yang digunakan dalam algoritma random forest dalam klasifikasi gangguan tidur antara lain:

- 1) Jumlah Pohon (*n\_estimators*): 10. Jumlah pohon yang relatif kecil ini dipilih untuk mengurangi waktu komputasi.
- 2) Kedalaman Maksimal Pohon (*max\_depth*): 20. Pembatasan ini mencegah pembentukan pohon yang terlalu dalam dan kompleks, yang bisa menyebabkan *overfitting*.
- 3) *Criterion = Gain Ratio*, digunakan sebagai kriteria pemisahan pada setiap node dalam pohon keputusan. Ini merupakan modifikasi dari *Information Gain* yang mempertimbangkan jumlah informasi yang berguna diperoleh dari pemisahan.

Parameter yang digunakan dalam penerapan algoritma Random Forest untuk klasifikasi gangguan tidur mencerminkan upaya untuk mencapai keseimbangan optimal antara kompleksitas model dan kemampuan generalisasi. Pengaturan parameter ini mencakup penggunaan jumlah pohon keputusan (*number of trees*) yang cukup untuk mengurangi variabilitas hasil, serta penerapan teknik seperti *pruning* dan pembatasan kedalaman pohon (*max depth*) guna mencegah *overfitting*, terutama pada data pelatihan. Selain itu, pemilihan fitur secara acak pada setiap pembentukan pohon juga berperan penting dalam meningkatkan keanekaragaman model dan memastikan kemampuan model untuk menangani data yang tidak seimbang atau mengandung *noise*. Dengan pendekatan ini, algoritma Random Forest mampu mempertahankan kinerja yang baik dalam memproses dataset kompleks seperti *Sleep Health and Lifestyle Dataset*. Akurasi yang diperoleh dalam klasifikasi gangguan tidur mencapai 89.69%, menunjukkan bahwa algoritma ini efektif dalam menangkap pola-pola signifikan yang terdapat dalam data.

b. K-Nearest Neighbors (KNN)

Dalam proses penerapan algoritma K-Nearest Neighbors (KNN) dalam klasifikasi gangguan tidur menggunakan aplikasi RapidMiner. Berikut adalah proses yang dilakukan menggunakan aplikasi tersebut:



Sumber: Hasil Penelitian (2024)

Gambar 3. Penerapan algoritma KNN menggunakan Rapid Miner

Konfigurasi parameter utama dari K-Nearest Neighbors (KNN) yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 4. Hasil eksperimen dengan mengubah nilai K

Nilai K	Hasil Akurasi
5	88.17%
6	88.17%
7	87.79%
8	87.40%

Hasil klasifikasi menggunakan algoritma K-Nearest Neighbors (KNN) menunjukkan bahwa akurasi terbaik yang diperoleh adalah sebesar 88.17% dengan nilai  $K = 5$ . Nilai  $K$  ini mengindikasikan bahwa model mempertimbangkan lima tetangga terdekat dalam menentukan kategori gangguan tidur untuk setiap data yang diuji. Performanya mencerminkan kemampuan KNN dalam memanfaatkan pola lokal dari data untuk membedakan kategori gangguan tidur, yaitu 'None,' 'Sleep Apnea,' dan 'Insomnia.'

Berdasarkan hasil penelitian, algoritma Random Forest memberikan akurasi yang lebih tinggi, yaitu 89,69%, dibandingkan dengan algoritma K-Nearest Neighbors (KNN) yang mencapai akurasi 88.17%. Hal ini menunjukkan bahwa Random Forest lebih unggul dalam mengklasifikasikan gangguan tidur pada dataset yang digunakan, memberikan performa yang lebih baik dalam menangkap pola dari data yang ada.

### 5. Evaluasi Model

Metode yang digunakan dalam proses evaluasi model adalah metode yang menghasilkan nilai akurasi terbaik yaitu Random Forest. *Confusion matrix* yang dihasilkan memberikan gambaran tentang kinerja model dalam mengklasifikasikan tiga kategori gangguan tidur: *None* (tidak ada gangguan), *Sleep Apnea*, dan *Insomnia*.

Tabel berikut menyajikan hasil detail dari *confusion matrix*:

Tabel 5. *Confusion Matrix*

	<i>True None</i>	<i>True Sleep Apnea</i>	<i>True Insomnia</i>	<i>Class Prediction</i>
<i>Pred. None</i>	147	14	6	88.02%
<i>Pred. Sleep Apnea</i>	2	41	1	93.18%
<i>Pred. Insomnia</i>	4	10	47	92.16%
<i>Class recall</i>	96.08%	74.55%	87.04%	

Sumber: Hasil Penelitian (2024)

Tabel 6. Penjabaran Nilai TP, TN, FP, FN dari masing-masing kelas

	TP	TN	FP	FN
None	147	88	6	20
Sleep Apnea	41	194	24	3
Insomnia	47	188	7	14

Sumber: Hasil Penelitian (2024)

Berikut untuk penjabaran dari nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) dari masing-masing kelas:

a. Kinerja Klasifikasi pada Kategori "None" (Tidak Ada Gangguan Tidur):

- 1) *True Positive* (TP): Model berhasil mengklasifikasikan 147 sampel dengan benar sebagai tidak memiliki gangguan tidur.

- 2) *True Negative* (TN): Model mengklasifikasikan 41 sampel Sleep Apnea diklasifikasikan sebagai Sleep Apnea dan 47 sampel Insomnia sebagai Insomnia.
  - 3) *False Positive* (FP): Model salah mengklasifikasikan 2 sampel yang seharusnya diklasifikasikan sebagai Sleep Apnea dan 4 sampel sebagai Insomnia.
  - 4) *False Negative* (FN): Model salah mengklasifikasikan 14 sampel yang seharusnya diklasifikasikan sebagai Sleep Apnea dan 6 sampel sebagai Insomnia.
  - 5) *Recall: Recall* untuk kategori ini adalah 96.08%, yang berarti model sangat baik dalam mendeteksi sampel yang benar-benar tidak memiliki gangguan tidur.
- b. Kinerja Klasifikasi pada Kategori "*Sleep Apnea*":
- 1) *True Positive* (TP): Model berhasil mengidentifikasi 41 sampel dengan benar sebagai Sleep Apnea.
  - 2) *True Negative* (TN): Sebanyak 147 sampel None diklasifikasikan sebagai "None", dan 194 sampel Insomnia sebagai Insomnia.
  - 3) *False Negative* (FN): Sebanyak 14 sampel Sleep Apnea salah diklasifikasikan sebagai "None", dan 10 sampel sebagai Insomnia.
  - 4) *False Positive* (FP): Terdapat 2 sampel yang salah diklasifikasikan sebagai Sleep Apnea padahal sebenarnya tidak memiliki gangguan tidur, serta 1 sampel yang diklasifikasikan sebagai Sleep Apnea padahal merupakan Insomnia.
  - 5) *Recall: Recall* untuk kategori ini adalah 74.55%. Meskipun recall cukup tinggi, ada kesalahan yang signifikan dalam mendeteksi Sleep Apnea, yang menunjukkan bahwa model memiliki kesulitan dalam membedakan Sleep Apnea dari kategori lainnya.
- c. Kinerja Klasifikasi pada Kategori "*Insomnia*":
- 1) *True Positive* (TP): Model berhasil mengklasifikasikan 47 sampel dengan benar sebagai Insomnia.
  - 2) *True Negative* (TN): Model mengklasifikasikan 147 sampel None sebagai "None" dan 41 sampel Sleep Apnea sebagai Sleep Apnea.
  - 3) *False Negative* (FN): Model salah mengklasifikasikan 6 sampel Insomnia sebagai "None" dan 1 sampel sebagai Sleep Apnea.
  - 4) *False Positive* (FP): Ada 4 sampel tanpa gangguan tidur dan 10 sampel dengan Sleep Apnea yang salah diklasifikasikan sebagai Insomnia.
  - 5) *Recall: Recall* untuk kategori Insomnia adalah 87.04%, menunjukkan bahwa model cukup baik dalam mendeteksi Insomnia, tetapi masih ada beberapa kesalahan.
- d. Keseluruhan Kinerja Model:
- 1) *Akurasi Prediksi*: Akurasi prediksi dalam klasifikasi masing-masing kategori (88.02% untuk None, 93.18% untuk Sleep Apnea, dan 92.16% untuk Insomnia) menunjukkan bahwa model bekerja cukup baik secara keseluruhan. Akurasi ini mencerminkan persentase prediksi yang benar untuk setiap kategori dibandingkan dengan total prediksi yang dibuat untuk kategori tersebut.
  - 2) *Imbalanced Data*: Meskipun model ini bekerja baik dalam mendeteksi kategori "None" dan "Insomnia", deteksi untuk kategori "Sleep Apnea" lebih sulit, yang mungkin menunjukkan adanya ketidakseimbangan data atau perbedaan karakteristik fitur yang signifikan antara kategori tersebut.

## **KESIMPULAN**

Kesimpulan dari penelitian ini menunjukkan bahwa penerapan algoritma Random Forest pada klasifikasi gangguan tidur, dengan menggunakan data kesehatan dan gaya hidup dari dataset Sleep Health and Lifestyle, berhasil menghasilkan model dengan akurasi yang signifikan, yaitu 89,69%. Algoritma ini terbukti efektif dalam mengklasifikasikan data ke dalam kategori 'None', 'Sleep Apnea', dan 'Insomnia', dengan performa terbaik tercatat pada kategori 'None', yang memiliki recall mencapai 96,08%. Namun, tantangan masih ditemukan dalam mengklasifikasikan kategori 'Sleep Apnea', di mana recall hanya mencapai 74,55%, mengindikasikan bahwa model ini mengalami kesulitan dalam mendeteksi gangguan tersebut dibandingkan dengan kategori lainnya. Hal ini mungkin disebabkan oleh ketidakseimbangan data atau perbedaan karakteristik fitur yang signifikan antara kategori-kategori tersebut.

Sebagai perbandingan, penelitian ini juga telah dilakukan menggunakan algoritma K-Nearest Neighbors (KNN) pada dataset yang sama, dengan hasil akurasi sebesar 87,02% untuk nilai  $K = 5$ . Walaupun hasil ini menunjukkan performa yang baik, algoritma Random Forest telah menunjukkan akurasi yang lebih tinggi dan lebih efektif dalam menangani data dengan banyak variabel.

Meskipun demikian, penggunaan algoritma Random Forest memberikan manfaat yang signifikan, terutama dalam hal menangani data dengan banyak variabel dan mengurangi risiko *overfitting*. Penelitian ini menggarisbawahi bahwa Random Forest merupakan alat yang andal untuk klasifikasi gangguan tidur, meskipun penyesuaian lebih lanjut diperlukan untuk meningkatkan performa dalam mendeteksi gangguan tidur tertentu, seperti Sleep Apnea.

Rencana penelitian selanjutnya akan mencakup penyeimbangan data untuk mengatasi ketidakseimbangan kategori dan eksplorasi algoritma lain seperti XGBoost atau LightGBM untuk membandingkan performa. Selain itu, penambahan fitur baru yang lebih relevan serta penerapan teknik hyperparameter tuning diharapkan dapat meningkatkan akurasi dan recall, khususnya untuk kategori Sleep Apnea. Penelitian lebih lanjut juga disarankan untuk fokus pada interpretasi model guna memahami lebih dalam faktor-faktor yang mempengaruhi deteksi masing-masing kategori gangguan tidur.

## REFERENSI

- Alvi Norma Utami. (2020). Klasifikasi Gangguan Tidur REM Behaviour Disorder Berdasarkan Sinyal EEG menggunakan Machine Learning. *Jurnal Sistem Cerdas*, 3(3), 216–230. <https://doi.org/10.37396/jsc.v3i3.68>
- Azmi, B. N., Hermawan, A., & Avianto, D. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(4), Article 4. <https://doi.org/10.35746/jtim.v4i4.298>
- Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). COVID-19 Image Data Collection: Prospective Predictions Are the Future. *Machine Learning for Biomedical Imaging*, 1(December 2020), 1–38. <https://doi.org/10.59275/j.melba.2020-48g7>
- Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). Multi-label Classifier Performance Evaluation with Confusion Matrix. *Computer Science & Information Technology*, 01–14. <https://doi.org/10.5121/csit.2020.100801>
- Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(9), Article 9.
- Madani, P. M. S., Rohana, T., Baihaqi, K. A., & Fauzi, A. (2024). Perbandingan Kinerja Klasifikasi Penyakit Ginjal Menggunakan Algoritma Support Vector Machine (SVM) dan Decision Tree (DT). *Building of Informatics, Technology and Science (BITS)*, 6(1), Article 1. <https://doi.org/10.47065/bits.v6i1.5206>
- Matur, Y. P., Simon, M. G., & Ndorang, T. A. (2021). HUBUNGAN KECANDUAN GAME ONLINE DENGAN KUALITAS TIDUR PADA REMAJA SMA NEGERI DI KOTA RUTENG. *Wawasan Kesehatan*, 6(2), Article 2.
- Maulida, D., & Nasution, Y. R. (2024). Prediksi Penjualan Produk Pepsodent Unilever dengan Algoritma K-Nearest Neighbor. *Journal of Computer System and Informatics (JoSYC)*, 5(4), Article 4. <https://doi.org/10.47065/josyc.v5i4.5718>
- Prabowo, I. A., Remawati, D., & Wardana, A. P. W. (2020). Klasifikasi Tingkat Gangguan Tidur Menggunakan Algoritma Naïve Bayes. *Jurnal Teknologi Informasi dan Komunikasi (TIKOMSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.519>
- Prasetyo, R. B. (2024). Prediksi Dini Penyakit Diabetes Pada Ibu Hamil Dengan Algoritma Random Forest. *JIMU: Jurnal Ilmiah Multidisipliner*, 2(04), Article 04. <https://doi.org/10.70294/jimu.v2i04.440>
- Putra, J. L., & Hidayat, W. F. (2024). Prediksi Kualitas Tidur: Pendekatan Machine Learning yang Mengintegrasikan Faktor Kesehatan dan Lingkungan. *Computer Science (CO-SCIENCE)*, 4(2), Article 2. <https://doi.org/10.31294/coscience.v4i2.4737>
- Rasmiati, R., Herawati, T., & Sukmarini, L. (2023). Terapi Komplementer dalam Mengatasi Gangguan Tidur Pasien yang Menjalani Hemodialisis. *Journal of Telenursing (JOTING)*, 5(1), Article 1. <https://doi.org/10.31539/joting.v5i1.6261>
- Sleep Health and Lifestyle Dataset*. (n.d.). Retrieved August 22, 2024, from <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>
- Sukmo, E. G. G., Munawar, Firmansyah, G., & Tjahjono, B. (2024). PERBANDINGAN PREDIKSI KEBERLANJUTAN POLIS ASURANSI DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES, KNN, DAN SVM. *Jurnal Bisnis Dan Manajemen*, 4(2), Article 2.

- Tang, S., Yuan, S., & Zhu, Y. (2020). Data Preprocessing Techniques in Convolutional Neural Network Based on Fault Diagnosis Towards Rotating Machinery. *IEEE Access*, 8, 149487–149496. IEEE Access. <https://doi.org/10.1109/ACCESS.2020.3012182>
- Widasari, E. R., Tanno, K., & Tamura, H. (2020). Automatic Sleep Disorders Classification Using Ensemble of Bagged Tree Based on Sleep Quality Features. *Electronics*, 9(3), Article 3. <https://doi.org/10.3390/electronics9030512>