

ANALISIS KOMPARASI KESTABILAN ALGORITMA DECISION TREE DAN RANDOM FOREST UNTUK KLASIFIKASI GANGGUAN TIDUR

Almunajat Amirul Soleh¹, Gusti Arviana Rahman^{*2}

^{1,2}Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Halu Oleo, Kendari

Email: ¹almunajatmirul@gmail.com, ²arviana.rahman@uho.ac.id
^{*}Penulis Korespondensi

Abstrak

Sleep disorders, juga dikenal sebagai gangguan tidur, merupakan masalah kesehatan serius yang berkorelasi erat dengan risiko hipertensi dan penyakit kardiovaskular. Mengingat mahalnya biaya diagnosis klinis konvensional, teknik data mining untuk deteksi dini menjadi alternatif yang efektif. Tujuan dari penelitian ini adalah untuk melakukan analisis perbandingan kinerja antara algoritma Decision Tree (C4.5) dan Random Forest untuk mengklasifikasikan gangguan tidur (None, Insomnia, dan Sleep Apnea). Untuk mengatasi ketidakseimbangan kelas, penelitian ini menggunakan dataset sekunder Sleep Health and Lifestyle. Untuk menguji stabilitas, metode validasi yang ketat, Stratified 10-Fold Cross-validation, digunakan. Hasil pengujian menunjukkan bahwa algoritma Random Forest mengalahkan Decision Tree dengan akurasi rata-rata 91,41% dan nilai Kappa 0,8473. Keunggulan utama algoritma Random Forest adalah kemampuan untuk meningkatkan sensitivitas deteksi kelas Insomnia secara signifikan menjadi 88,31%. Menurut analisis nilai fitur, tekanan darah diastolik dan kategori BMI (Overweight) adalah fitur yang paling umum dalam diagnosis. Random Forest dianggap lebih akurat dan stabil untuk digunakan sebagai model sistem pendukung keputusan medis.

Kata kunci: *data mining, gangguan tidur, klasifikasi, decision tree, random forest, stratified cross-validation*

Abstract

Sleep disorders are serious health conditions that are closely associated with an increased risk of hypertension and cardiovascular disease. Considering the high cost of conventional clinical diagnostic procedures, data mining techniques for early detection offer an effective alternative. This study aims to conduct a comparative performance analysis between the Decision Tree (C4.5) and Random Forest algorithms in classifying sleep disorders (None, Insomnia, and Sleep Apnea). To address class imbalance issues, this research utilizes the secondary Sleep Health and Lifestyle dataset. To evaluate model stability, a rigorous validation method, Stratified 10-Fold Cross-Validation, is employed. The experimental results indicate that the Random Forest algorithm outperforms the Decision Tree, achieving an average accuracy of 91.41% and a Kappa value of 0.8473. The primary advantage of the Random Forest algorithm lies in its ability to significantly improve the detection sensitivity of the Insomnia class to 88.31%. Based on feature importance analysis, diastolic blood pressure and the BMI category (Overweight) are identified as the most influential features in the diagnostic process. Random Forest is therefore considered a more accurate and stable model for medical decision support systems.

Keywords: *data mining, sleep disorder, classification, decision tree, random forest, stratified cross-validation*

1. PENDAHULUAN

Gangguan tidur merupakan masalah kesehatan global yang sering kali tidak teridentifikasi dengan cepat, meskipun memiliki keterkaitan yang kuat dengan berbagai penyakit degeneratif serius. Berbagai studi klinis menunjukkan bahwa gangguan pernapasan terkait tidur, khususnya Obstructive Sleep Apnea (OSA), merupakan faktor risiko independen terhadap hipertensi dan penyakit kardiovaskular. Sebuah tinjauan sistematis terbaru yang dipublikasikan pada tahun 2024 di jurnal medis terkemuka menunjukkan bahwa mekanisme hipoksia

periodik dan aktivasi sistem saraf simpatik pada individu dengan OSA secara langsung berkontribusi terhadap peningkatan tekanan darah dan risiko penyakit jantung [1]. Penelitian Antariksa (2021) mendukung hal tersebut dengan menunjukkan bahwa tingkat hipertensi meningkat seiring bertambahnya indeks Apnea Hypopnea Index (AHI), di mana peningkatan resistensi pembuluh darah menjadi faktor utama penyebab komplikasi kardiovaskular pada pasien yang tidak mendapatkan perawatan yang tepat [2]. Sayangnya, metode diagnosis yang diakui sebagai yang terbaik, seperti polisomnografi,

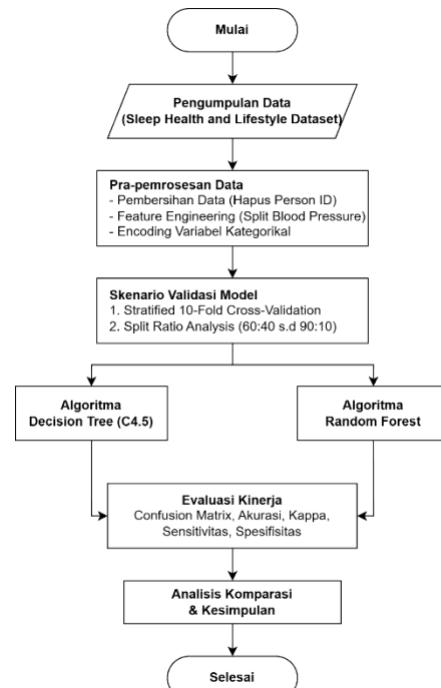
memerlukan biaya yang tinggi dan prosedur yang rumit. Oleh karena itu, penggunaan teknik Data Mining pada data kesehatan rutin menjadi solusi alternatif yang efisien untuk mendeteksi masalah sejak dini.

Penggunaan algoritma Machine Learning dalam klasifikasi medis telah berkembang pesat untuk mendukung sistem pengambilan keputusan klinis. Decision Tree (C4.5) dan Random Forest merupakan dua algoritma yang paling umum dibandingkan kinerjanya dalam bidang ini. Decision Tree memiliki keunggulan karena mudah dipahami, tetapi juga memiliki kelemahan yaitu bersifat tidak stabil, di mana perubahan kecil pada data dapat menyebabkan perubahan besar pada struktur pohon. Di sisi lain, metode ensemble yang disebut Random Forest lebih baik dalam melakukan generalisasi. Maulidiyyah et al. (2024) menunjukkan keunggulan Random Forest dalam penelitian klasifikasi diabetes, yang mengindikasikan efektivitas yang lebih tinggi serta stabilitas yang lebih baik dibandingkan dengan Decision Tree [3]. Hidayat, Sunyoto, dan Al Fatta (2023) melaporkan hasil serupa dalam klasifikasi penyakit jantung, yang menunjukkan bahwa metode Random Forest lebih efektif dalam menangani kompleksitas data medis dibandingkan dengan pohon keputusan tunggal [4].

Meskipun algoritma Random Forest telah terbukti efektif, penerapannya pada dataset tertentu, yaitu Sleep Health and Lifestyle, masih menghadapi tantangan terkait validitas metode pengujian. Sebuah penelitian sebelumnya oleh I. A. Hidayat (2023) menggunakan dataset yang sama untuk menerapkan algoritma Random Forest dalam klasifikasi gangguan tidur [5]. Namun, penelitian tersebut hanya menggunakan metode evaluasi standar tanpa validasi silang bertingkat, yang berpotensi menghasilkan model yang bias tergantung pada cara pembagian data. Metode validasi sederhana ini tidak cukup untuk mengukur stabilitas model secara objektif, terutama karena sampel data medis sering kali terbatas. Penelitian ini bertujuan untuk melakukan analisis perbandingan yang lebih mendalam antara algoritma Decision Tree dan Random Forest dengan mengatasi kekurangan yang diidentifikasi pada penelitian sebelumnya melalui pendekatan metodologis yang lebih ketat. Inovasi dari penelitian ini terletak pada penerapan teknik Stratified 10-Fold Cross-Validation yang bertujuan untuk memastikan keseimbangan proporsi kelas pada setiap pengujian, serta melakukan Analisis Sensitivitas Pembagian Data menggunakan metode Repeated Random Sub-sampling (Monte Carlo). Metode ini digunakan untuk menguji seberapa stabil model ketika rasio data pelatihan berubah dan seberapa baik model tersebut dapat menyeimbangkan sensitivitas deteksi antar kategori penyakit yang berbeda.

2. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif eksperimental untuk menganalisis kinerja algoritma Machine Learning dalam klasifikasi. Metodologi penelitian disusun secara tersusun berdasarkan kerangka kerja Knowledge Discovery in Database (KDD), dimulai dengan pengambilan data, dilanjutkan dengan prapemrosesan, validasi model, dan akhirnya evaluasi kinerja. Langkah-langkah rinci dari penelitian ini ditampilkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Data sekunder publik yang mencakup komponen kesehatan tidur dan gaya hidup digunakan dalam penelitian ini. Informasi demografi, tanda-tanda vital, dan kebiasaan sehari-hari hanyalah tiga belas fitur dari 374 catatan data dalam kumpulan data ini. Tujuan klasifikasi variabel adalah gangguan tidur yang terbagi menjadi tiga kategori diagnosis: Tidak Ada (Normal/Sehat), Insomnia, dan Sleep Apnea.

Sebelum pemodelan, kualitas data ditingkatkan melalui tahap pra-pemrosesan, atau preprocessing. Attribute identitas individu dihapus untuk mencegah kebocoran data, yang dapat menyebabkan kesalahan penilaian. Tekanan darah, atribut tekanan darah, diubah menjadi fitur melalui rekayasa fitur. Dua variabel numerik dalam karakteristik ini adalah tekanan darah diastolik dan sistol. Ada perbedaan klinis antara variabel-variabel ini mengenai potensi bahaya bagi jantung dan paru-paru. Akibatnya, variabel-variabel ini dipisahkan. Selain itu, teknik encoding diterapkan untuk mengubah data sehingga algoritma dapat memprosesnya secara matematis. Untuk mempertahankan makna tingkat risiko kesehatan, variabel kategori BMI, seperti jenis kelamin, pekerjaan, dan kategori BMI, diurutkan secara ordinal (Normal, Overweight, dan Obese).

Untuk memastikan keandalan hasil, model divalidasi dengan dua skenario pengujian bertingkat. Dalam skenario pertama, Stratified 10-Fold Cross-validation digunakan untuk melakukan evaluasi kinerja utama. Metode ini membagi dataset menjadi sepuluh bagian. Selanjutnya, model dilatih dan diuji secara bergantian sepuluh kali. Metode stratifikasi digunakan untuk memastikan bahwa rasio kelas Insomnia, Sleep Apnea, dan None tetap seimbang di setiap lipatan tes. Metode ini telah terbukti secara empiris dapat menghasilkan evaluasi akurasi yang lebih adil dan konsisten daripada validasi silang konvensional. Ini terutama berlaku untuk dataset medis dengan distribusi kelas yang tidak merata [6]. Uji Stabilitas (Sensitivity Analysis) adalah skenario kedua yang menggunakan metode sub-sampling acak berulang (Monte Carlo). Dataset dibagi menjadi empat tingkat rasio data latih dan data uji, masing-masing 60:40 dan 90:10. Untuk mengevaluasi konsistensi akurasi model terhadap variabilitas data, pengujian dilakukan sebanyak lima kali iterasi pada setiap rasio dengan pengacakan data yang berbeda.

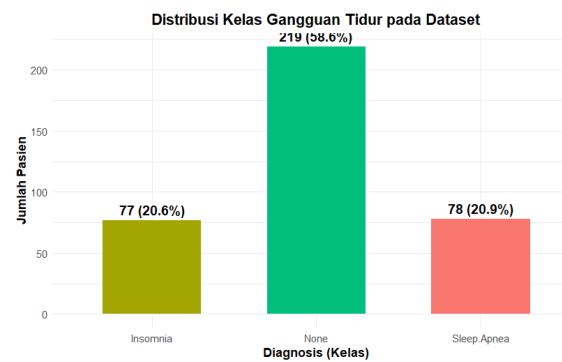
Dua algoritma digunakan untuk pemodelan klasifikasi: Random Forest (C4.5) dan Decision Tree. Dalam algoritma Random Forest, parameter jumlah pohon (ntree) diatur sebanyak 500 pohon untuk mencapai konvergensi kesalahan yang optimal melalui mekanisme bagging dan fitur kebetulan. Untuk mengevaluasi kinerja kedua model, matriks kebingungan (Confusion Matrix) digunakan untuk menghitung metrik Akurasi, Presisi, Recall (Sensitivitas), dan F1-Score. Fokus evaluasi terletak pada kemampuan model untuk menyeimbangkan sensitivitas deteksi antar kelas penyakit dan mempertahankan stabilitas performa dalam kondisi data latih yang terbatas.

3. HASIL DAN PEMBAHASAN

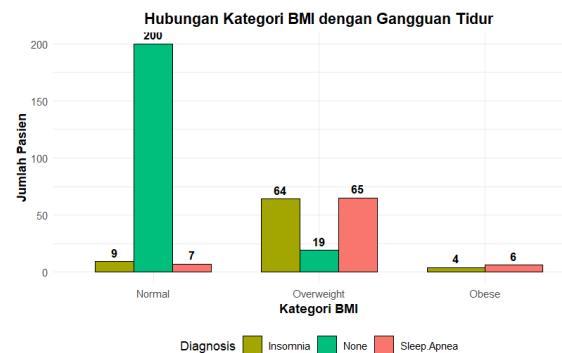
3.1. Analisis Deskriptif dan Korelasi Data

Sebelum pemodelan dilakukan, tahap awal penelitian melibatkan analisis data untuk memahami karakteristik distribusi dataset. Gambar 2 menunjukkan bahwa analisis pertama berkonsentrasi pada distribusi variabel target, yang dikenal sebagai gangguan tidur. Data menunjukkan ketidakseimbangan kelas yang signifikan. Dalam 219 sampel, kategori None (Normal) menyumbang 58,6% dari populasi dataset. Namun, proporsi penderita insomnia 20,6% (77 sampel) dan insomnia 20,9% (78 sampel) hampir sama. Untuk menjaga proporsi kelas yang representatif, ketidakseimbangan ini menjadi dasar penerapan metode cross-validation stratified di tahap berikutnya.

Gambar 2. Distribusi Frekuensi Kelas Gangguan Tidur

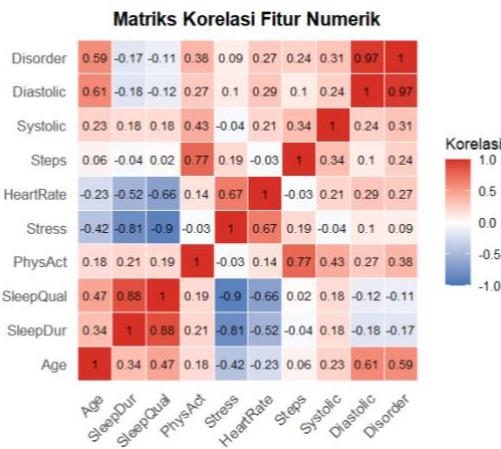


Analisis tambahan dilakukan untuk menentukan hubungan antara kategori indeks massa tubuh (BMI) dan diagnosis penyakit. Gambar 3 menunjukkan temuan medis penting: sebagian besar orang dengan berat badan normal cenderung sehat, dan 200 dari 216 sampel memiliki diagnosis None. Di sisi lain, kelompok orang gemuk sangat rentan terhadap gangguan tidur, dengan 64 sampel mengalami insomnia dan 65 sampel mengalami gangguan tidur. Pola visual ini menunjukkan bahwa peningkatan berat badan adalah faktor risiko utama yang membedakan individu yang sehat dari individu yang menderita gangguan tidur.



Gambar 3. Hubungan Kategori BMI dengan Diagnosis Gangguan Tidur

Untuk menemukan kekuatan hubungan linear antar variabel, analisis korelasi antar atribut numerik juga dilakukan. Hubungan linier positif yang signifikan ditemukan antara variabel tekanan darah diastolik dan label gangguan tidur, dengan nilai koefisien 0,97. Ini adalah bukti statistik yang signifikan, seperti yang ditunjukkan dalam matriks korelasi yang ditunjukkan pada Gambar 4.

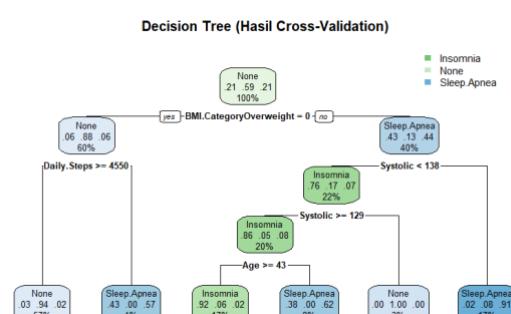


Gambar 4. Matriks Korelasi Antar Variabel Numerik

Selain tekanan darah, variabel usia menunjukkan korelasi positif moderat sebesar 0,59, menunjukkan bahwa risiko gangguan tidur meningkat seiring bertambahnya usia. Salah satu temuan yang menarik adalah hubungan negatif yang sangat kuat sebesar -0,90 antara tingkat stres (tingkat stres) dan kualitas tidur (kualitas tidur). Secara keseluruhan, hasil dari analisis deskriptif dan korelasi ini memvalidasi relevansi atribut klinis yang digunakan. Hasil-hasil ini juga memberikan dasar yang kuat untuk langkah selanjutnya dalam pelatihan model klasifikasi.

3.2. Evaluasi Kinerja Klasifikasi

Untuk memulai tahap pemodelan klasifikasi, algoritma Decision Tree (C4.5) digunakan untuk membuat aturan keputusan. Struktur pohon yang dihasilkan menempatkan simpul akar (root node) sebagai atribut kategori BMI, khususnya kategori obesitas, seperti yang ditunjukkan pada Gambar 5. Hal ini selaras dengan hasil dari tahap eksplorasi data, di mana berat badan adalah faktor utama yang membedakan antara pasien sehat dan sakit. Faktor Tekanan Darah Sistolik dan Usia digunakan sebagai penentu akhir diagnosis setelah percabangan lebih lanjut pada pohon keputusan. Secara khusus, aturan keputusan menunjukkan bahwa orang overweight dengan tekanan darah sistolik di atas 138 mmHg sangat mungkin menderita gangguan tidur.



Gambar 5. Struktur Pohon Keputusan (Decision Tree)

Meskipun Decision Tree memiliki keunggulan dalam hal interpretabilitas visual, evaluasi kinerja statistik menggunakan cross-validation Stratified 10-Fold menunjukkan bahwa algoritma Random Forest lebih baik dalam prediksi. Pada semua metrik evaluasi penting, algoritma Random Forest secara konsisten mengungguli Decision Tree, seperti yang disajikan dalam Tabel 1. Decision Tree memiliki akurasi 89,47% dengan nilai Kappa 0,8133, sedangkan Random Forest memiliki akurasi rata-rata sebesar 91,41%. Nilai Kappa di atas 0,80 pada kedua model menunjukkan bahwa tingkat kesepakatan klasifikasi yang dihasilkan termasuk dalam kategori "hampir sempurna" (almost perfect agreement), tetapi Random Forest terbukti lebih baik dalam mengurangi kesalahan prediksi.

Tabel 1. Perbandingan Performa Algoritma (Rata-rata 10-Fold CV)

| Metrik Evaluasi | Decision Tree | Random Forest |
|----------------------------|---------------|---------------|
| Akurasi | 89,47% | 91,41% |
| Kappa | 0,8133 | 0,8473 |
| Sensitivitas (Insomnia) | 77,92% | 88,31% |
| Sensitivitas (Sleep Apnea) | 92,31% | 89,74% |
| Metrik Evaluasi | Decision Tree | Random Forest |

Gambar 6 menunjukkan confusion matrix yang dianalisis untuk mendapatkan pemahaman yang lebih rinci tentang distribusi kesalahan. Salah satu kelemahan utama algoritma Decision Tree adalah kemampuannya untuk menemukan kelas insomnia, dengan sensitivitas hanya 77,92%. Ini menunjukkan bahwa sejumlah besar pasien yang menderita insomnia salah dianggap sehat atau menderita apnea tidur.

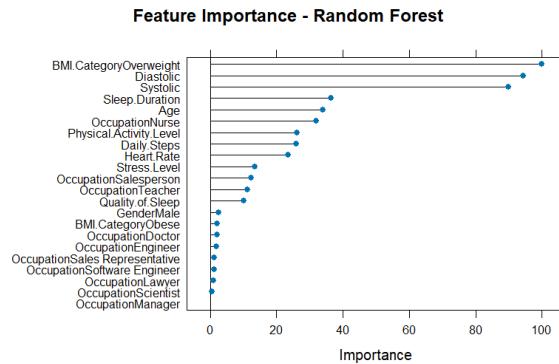


Gambar 6. Perbandingan Confusion Matrix: (a) Decision Tree vs (b) Random Forest

Namun, kelemahan ini secara signifikan diperbaiki oleh algoritma Random Forest. Dengan menggabungkan prediksi dari berbagai pohon keputusan, mekanisme ensemble mampu meningkatkan sensitivitas deteksi kelas Insomnia menjadi 88,31% tanpa mengurangi akurasi kelas lainnya. Peningkatan kinerja ini menunjukkan bahwa Random Forest lebih baik dalam menangkap pola data yang kompleks dan mengurangi varians kesalahan yang sering terjadi pada pohon keputusan tunggal. Akibatnya, Random Forest adalah model yang lebih cocok untuk diagnosis klinis pada dataset ini.

3.3. Analisis Variabel Dominan

Untuk memvalidasi relevansi medis dari model yang dibangun, dilakukan ekstraksi nilai Mean Decrease Gini atau peringkat kepentingan fitur (Feature Importance) dari algoritma Random Forest. Hasil pemeringkatan variabel disajikan pada Gambar 7.



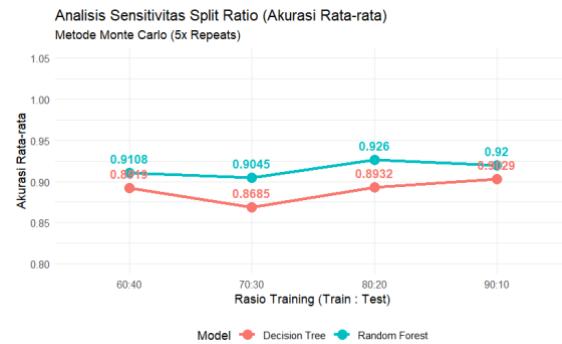
Gambar 7. Peringkat Kepentingan Fitur pada Random Forest

Grafik tersebut menunjukkan bahwa kategori BMI (overweight) memiliki skor kepentingan tertinggi dengan 100 persen, diikuti oleh tekanan darah diastolik 94,32 dan sistolik 89,96. Karakteristik gaya hidup seperti durasi tidur (sleep duration) dan tingkat aktivitas fisik berada di urutan menengah.

Hasil ini sejalan dengan penelitian baru-baru ini oleh Putri et al. (2025) yang melihat fitur klinis dari sleep apnea dan menemukan bahwa indeks massa tubuh dan tekanan darah adalah dua faktor yang paling penting untuk meningkatkan akurasi deteksi [7]. Hal serupa juga dikonfirmasi oleh penelitian Fernandez et al. (2025), yang menemukan bahwa obesitas dan tekanan darah tinggi merupakan faktor yang paling signifikan dalam klasifikasi gangguan tidur berbasis gaya hidup [8]. Fakta bahwa model Random Forest tidak hanya akurat secara statistik tetapi juga relevan secara klinis ditunjukkan oleh konsistensi antara temuan pentingnya fitur model dengan literatur medis terbaru. Ini menunjukkan bahwa pengendalian berat badan dan tekanan darah pasien harus menjadi fokus utama intervensi medis untuk gangguan tidur.

3.4. Uji Stabilitas Model

Tujuan dari pengujian terakhir adalah untuk mengukur tingkat stabilitas model—atau ketahanan—terhadap perubahan ketersediaan data. Metode Repeated Random Subsampling (Monte Carlo) digunakan untuk melakukan uji ini sebanyak lima kali iterasi pada berbagai variasi rasio pembagian data uji dan data latih. Gambar 8 menunjukkan tren rata-rata akurasi kedua model.



Gambar 8. Tren Stabilitas Akurasi pada Berbagai Rasio Data

Berdasarkan grafik di atas, terlihat pola kesenjangan performa yang semakin melebar seiring berkurangnya jumlah data latih. Rincian statistik performa disajikan dalam **Tabel 2**.

Tabel 2. Rekapitulasi Kinerja Model pada Berbagai Rasio (Rata-rata 5 Iterasi)

| Rasio Data | Algoritma | Akurasi | Presisi | Recall | F1-Score |
|------------|---------------|---------|---------|--------|----------|
| 60:40 | Decision Tree | 0.8919 | 0.8734 | 0.8665 | 0.8688 |
| | Random Forest | 0.9108 | 0.8909 | 0.8854 | 0.8867 |
| 70:30 | Decision Tree | 0.8685 | 0.8414 | 0.8483 | 0.8436 |
| | Random Forest | 0.9045 | 0.8875 | 0.8801 | 0.8819 |
| 80:20 | Decision Tree | 0.8932 | 0.876 | 0.873 | 0.873 |
| | Random Forest | 0.926 | 0.9107 | 0.9089 | 0.909 |
| 90:10 | Decision Tree | 0.9029 | 0.9083 | 0.8571 | 0.8621 |
| | Random Forest | 0.92 | 0.9094 | 0.8794 | 0.8846 |

Algoritma Random Forest memiliki ketahanan yang lebih baik, seperti yang ditunjukkan oleh data di Tabel 2. Random Forest masih mampu mempertahankan akurasi rata-rata 91,08% dengan F1-Score 0,8867 pada kondisi data latih yang minimal (rasio 60:40). Sebaliknya, Decision Tree mengalami variasi yang lebih nyata dalam performa, terutama pada rasio 70:30, di mana akurasi turun menjadi 86,85%. Ini menunjukkan bahwa metode ensemble efektif dalam mengurangi varians dan risiko otentikasi.

4. KESIMPULAN

Berdasarkan hasil pengujian, algoritma Random Forest menunjukkan kinerja klasifikasi yang lebih baik daripada Decision Tree (C4.5) dalam diagnosis gangguan tidur. Dengan menggunakan metode cross-validation stratified 10-fold, evaluasi menunjukkan bahwa Random Forest mencapai akurasi rata-rata sebesar 91,41% dengan nilai Kappa 0,8473. Keunggulan paling signifikan dari Random Forest terletak pada kemampuan untuk meningkatkan sensitivitas deteksi kelas Insomnia menjadi 88,31%, mengatasi kelemahan Decision Tree sebelumnya,

yang memiliki tingkat kesalahan prediksi yang cukup tinggi pada kelas Insomnia.

Tekanan darah diastolik dan kategori BMI (berlebihan berat badan) adalah komponen fisiologis yang memengaruhi risiko gangguan tidur, menurut penelitian ini. Ini dipastikan melalui analisis korelasi, struktur pohon keputusan, dan peringkat kepentingan fitur (Feature Importance). Hasil uji stabilitas metode Monte Carlo menunjukkan bahwa Random Forest adalah algoritma yang kuat dan tangguh. Model ini dapat mempertahankan akurasi dan stabilitas yang tinggi meskipun diuji pada jumlah data latih yang sangat kecil. Akibatnya, Random Forest diusulkan sebagai strategi utama untuk membangun sistem yang membantu keputusan klinis dalam menemukan gangguan tidur dini.

5. DAFTAR PUSTAKA

- [1] S. Amen *et al.*, "Obstructive Sleep Apnea and Cardiovascular Diseases: A Systematic Review and Meta-Analysis of Prospective Studies," *Cureus*, vol. 16, no. 10, Oct. 2024.
- [2] B. Antariksa, "Obstructive Sleep Apnea (OSA) dan Penyakit Jantung," *Jurnal Respirologi Indonesia*, vol. 41, no. 1, pp. 1-9, Jan. 2021.
- [3] N. A. Maulidiyyah, Trimono, A. T. Damaliana, and D. A. Prasetya, "Comparison of Decision Tree and Random Forest Methods in the Classification of Diabetes Mellitus," *JIKO (Journal of Informatics and Computers)*, vol. 7, no. 2, pp. 79-87, Aug. 2024.
- [4] H. Hidayat, A. Sunyoto, and H. Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 31-40, 2023.
- [5] I. A. Hidayat, "Classification of Sleep Disorders Using Random Forest on Sleep Health and Lifestyle Dataset," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 2, pp. 71-76, 2023.
- [6] S. Widodo, H. Brawijaya, and S. Samudi, "Stratified K-fold Cross Validation Optimization on Machine Learning for Prediction," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 6, no. 4, pp. 2407-2414, 2022.
- [7] T. R. Putri *et al.*, "Application of Support Vector Machine Algorithm For Classification of Sleep Disorders," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 5, no. 1, Oct. 2025.
- [8] S. Fernandez, A. Riyandi, S. Wijayanto, and S. Sukmadiningtyas, "Classification Of Sleep Disorders Based on Lifestyle and Health Factors Using Random Forest and HistGradientBoosting," *Journal of Artificial Intelligence and Software Engineering*, vol. 5, no. 2, pp. 783-789, 2025.
- [9] M. Y. Iskandar and H. W. Nugroho, "Comparative Evaluation of Decision Tree and Random Forest for Lung Cancer Prediction Based on Computational Efficiency and Predictive Accuracy," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3392-3404, 2025.
- [10] L. Tharmalingam, "Sleep Health and Lifestyle Dataset," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>