

# Klasifikasi Gangguan Tidur Menggunakan Metode Decision Tree dan Algoritma Genetika

Wisti Dwi Septiani

Universitas Bina Sarana Informatika, Indonesia

[wisti.wst@bsi.ac.id](mailto:wisti.wst@bsi.ac.id)

## ABSTRAK

Pola tidur yang baik berperan penting dalam menjaga kesehatan tubuh serta keseimbangan fisik dan mental seseorang. Gaya hidup yang sehat seperti pola makan, aktivitas fisik dan kebiasaan pola hidup sehat berperan penting dalam mempengaruhi kualitas tidur. Gangguan tidur merupakan masalah kesehatan yang semakin banyak dialami oleh individu global saat ini dan berpotensi memberikan dampak negatif terhadap kualitas hidup mereka. Dalam era kecerdasan buatan, analisis data menjadi salah satu pendekatan yang efektif untuk memahami pola dan prediksi terkait kesehatan. Tujuan penelitian ini adalah memprediksi gangguan tidur berdasarkan kualitas tidur dan gaya hidup, menggunakan algoritma klasifikasi data mining. Metode yang digunakan dalam penelitian ini adalah implementasi metode Decision Tree, dengan seleksi fitur menggunakan Algoritma Genetika untuk meningkatkan akurasi metode. Hasil penelitian ini adalah kenaikan nilai akurasi sebesar 10,72% dengan pengujian optimasi menggunakan metode Algoritma Genetika, juga dihasilkan pohon keputusan dari metode Decision Tree. Kesimpulan dari penelitian ini adalah metode optimasi Algoritma Genetika dapat membantu meningkatkan kehandalan dan rule yang dihasilkan dari pohon keputusan Metode Decision Tree dipilih karena mampu memberikan prediksi yang akurat dan menghasilkan rule dari pohon keputusan yang mudah dipahami dan dapat membantu kita untuk memprediksi gangguan tidur sehingga dapat meningkatkan kesadaran akan pentingnya hidup sehat dengan menjaga kualitas tidur dan gaya hidup yang sehat. Hasil penelitian ini juga menunjukkan bahwa durasi tidur, tingkat stres, aktivitas fisik, umur, dan BMI memberikan pengaruh signifikan terhadap kualitas tidur seseorang.

**Kata Kunci:** Decision Tree, Algoritma Genetika, Data Mining, Metode Klasifikasi, Gangguan Tidur

## PENDAHULUAN

Gangguan tidur merupakan masalah kesehatan yang semakin banyak dialami oleh masyarakat modern. Pola tidur yang tidak teratur, stres yang tinggi, kurangnya aktivitas fisik, dan faktor gaya hidup lainnya menjadi penyebab utama gangguan tidur. Kondisi ini tidak hanya mempengaruhi kualitas hidup individu, tetapi juga berkontribusi terhadap risiko penyakit kronis seperti obesitas, diabetes, hipertensi, gangguan kardiovaskular dan gangguan mental. Pada era modern, pola tidur Masyarakat kerap terganggu oleh berbagai faktor gaya hidup, seperti stress pekerjaan, paparan cahaya dari perangkat elektronik, konsumsi kafein berlebihan, dan pola makan yang tidak teratur (Nawawi & Fatah, 2024). Oleh karena itu, penting untuk memahami faktor-faktor yang mempengaruhi gangguan tidur agar dapat mencegah dan mengatasinya secara efektif. Gangguan tidur pada dataset penelitian ini terdiri dari insomnia, sleep apnea dan kondisi tidak ada gangguan tidur. Insomnia kondisi gangguan tidur dimana seseorang kesulitan untuk tidur, sleep apnea Dimana kondisi klinis ditandai dengan adanya kesulitan bernafas saat tidur, sehingga bisa menurunkan kualitas tidur seseorang (Sari, 2024). Pada penelitian ini Dalam era kecerdasan buatan, analisis data menjadi salah satu pendekatan yang efektif untuk memahami pola dan prediksi terkait kesehatan. Dataset Sleep Health and Lifestyle menyediakan informasi penting tentang durasi tidur,

tingkat stres, aktivitas fisik, dan indeks massa tubuh (BMI) yang dapat digunakan untuk menganalisis kemungkinan seseorang mengalami gangguan tidur.

Literatur studi tentang penelitian ini yaitu prediksi gangguan tidur menggunakan Support Vector Machine dengan akurasi 90,1% dan Neural Network dengan akurasi 91,2% yang menunjukkan kedua algoritma memiliki Tingkat akurasi yang memuaskan (Sari, 2024). Lalu penelitian Implementasi Algoritma K-Means Clustering untuk penentuan gangguan tidur dengan hasil pengujian silhouette score pada data Heart Rate, dan Daily Steps dapat disimpulkan bahwa pemilihan jumlah cluster sebanyak 3 menghasilkan skor sebesar 0.7233444078004578 menunjukkan struktur yang kuat (*strong structure*) (Maulida et al., 2024).

Pada penelitian ini dilakukan optimasi terhadap Algoritma Decision Tree menggunakan algoritma Genetika. Optimasi adalah proses menyelesaikan suatu masalah tertentu supaya berada pada kondisi yang paling menguntungkan dari suatu sudut pandang, yaitu berhubungan dengan pencarian nilai minimum atau nilai maksimum (Buani, 2016). Pada penerapan optimasi menggunakan algoritma genetika terdapat peningkatan akurasi dari 97,66% menjadi 99,33% pada algoritma Naïve Bayes untuk prediksi *fertility* (Buani, 2016) dan peningkatan akurasi dari 83,81% menjadi 86,47% pada algoritma C4.5 untuk prediksi *phising website* (Sunge, 2018). Decision Tree merupakan salah satu algoritma klasifikasi data mining. Tujuan dari penelitian ini adalah melakukan optimasi algoritma Decision Tree menggunakan Algoritma Genetika untuk meningkatkan nilai akurasinya sehingga prediksi yang dihasilkan lebih baik dan akurat.

## TINJAUAN PUSTAKA

### Data Mining

Proses menggali nilai tambahan dari kumpulan data yang sebelumnya tidak diketahui dikenal sebagai data mining (Sukri & Handrianto, 2024). Data Mining bertujuan utama untuk mengolah data di dalam basis data sehingga mendapatkan wawasan baru yang bermanfaat (Maulida et al., 2024). Salah satu metode terkenal dalam data mining adalah Algoritma C4.5. Algoritma ini digunakan untuk prediksi menggunakan teknik pohon keputusan dan dapat mengolah data numerik dan diskrit. Algoritma C4.5 juga memiliki kemampuan untuk menangani nilai atribut yang hilang dan menghasilkan aturan yang mudah dipahami (Sukri & Handrianto, 2024).

Data Mining juga dapat diartikan sebagai proses menganalisis data untuk menemukan dan merangkum informasi yang tidak diinginkan dan dapat dimengerti. Data mining juga sebuah proses mengekstraksi informasi dan pola dari sejumlah besar data tersebut. Data mining mencakup dari pengumpulan data, ekstraksi data, statistik data dan analisis data (Sari, 2024), lalu definisi yang lain menyatakan bahwa Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Nawawi & Fatah, 2024).

Data mining melibatkan kecerdasan buatan untuk mengidentifikasi pola yang dapat digunakan untuk membuat prediksi tentang perilaku masa depan. Analisis data mining melebihi apa yang dilakukan oleh sistem pendukung lainnya. Data mining tidak membutuhkan banyak waktu dan biaya yang murah. Data mining mencari informasi untuk memprediksi apa yang mungkin dilupakan atau tersembunyi pada data itu sendiri (Sari, 2024).

### Klasifikasi

Berdasarkan tugasnya, *data mining* dikelompokkan menjadi 6 yaitu deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi (Larose, 2005). Klasifikasi (taksonomi) adalah proses menempatkan objek tertentu (konsep) dalam satu set kategori, berdasarkan masing-masing objek (konsep) *property* (Gorunescu, 2011). Proses klasifikasi didasarkan pada empat komponen mendasar yaitu kelas, prediktor, *training set*, dan pengujian *dataset*.

Diantara model klasifikasi yang paling populer adalah *Decision/Classification Trees*, *Bayesian Classifiers/Naïve Bayes Classifiers*, *Neural Networks*, *Statistical Analysis*, *Genetic Algorithms*, *Rough Sets*, *K-Nearest Neighbor Classifier*, *Rule-based Methods*, *Memory Based Reasoning*, *Support Vector Machines* (Gorunescu, 2011). Tujuan dari klasifikasi adalah menemukan model yang dapat mengkategorikan atribut kedalam kategori atau kelas yang sesuai. Tujuan dari metode ini adalah untuk menentukan kategori dari objek yang modelnya belum diketahui. ID3, CART, dan C4.5 adalah beberapa teknik klasifikasi (Sukri & Handrianto, 2024)

### Decision Tree

Pohon Keputusan adalah hasil dari Algoritma C4.5, yang merupakan salah satu Algoritma klasifikasi yang sangat efektif dan sering digunakan dalam berbagai konteks. Algoritma C4.5 adalah pengembangan dari Algoritma Iterative Dichotomiser 3 (ID3) yang diperkenalkan oleh J. Ross Quinlan pada tahun 1996. Algoritma C4.5 menghasilkan sistem pembentukan pohon keputusan yang berhasil berkat perbaikan dari algoritma ID3. Beberapa perbaikan termasuk pemangkasan pohon, menangani hilangnya data, dan menangani data kontinu (Sukri & Handrianto, 2024)

Decision Tree merupakan alah satu algoritma dalam metode klasifikasi data mining yaitu Decision Tree yang menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal (Larose, 2005). Tahapan dalam membuat pohon keputusan dengan Decision Tree atau algoritma C4.5 (Gorunescu, 2011) yaitu:

1. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing atribut atau berdasarkan nilai *index entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai *index entropy*, dengan rumus:

$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

3. Hitung nilai gain dengan rumus:

$$Entropy\ split = \sum_{i=1}^p \binom{n-1}{n} \cdot IE(i)$$

4. Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat:
  - a. Semua tupel dalam *record* dalam simpul N mendapat kelas yang sama.
  - b. Tidak ada atribut dalam *record* yang dipartisi lagi.
  - c. Tidak ada *record* di dalam cabang yang kosong

### Algoritma Genetika

Menurut Zukhi dalam (Buani, 2016) Optimasi adalah proses menyelesaikan suatu masalah tertentu supaya berada pada kondisi yang paling menguntungkan dari suatu sudut pandang. Masalah yang harus diselesaikan berkaitan erat dengan data-data yang dapat dinyatakan dalam satu atau beberapa variabel. Pengertian menguntungkan biasanya berhubungan dengan pencarian nilai minimum atau pencarian nilai maksimum, bergantung pada sudut pandang yang digunakan. Algoritma Genetika merupakan suatu teknik pencarian (Search Technique) dan teknik optimasi yang cara kerjanya meniru proses utama dan perubahan struktur genetik pada makhluk hidup. Algoritma Genetika menggunakan probabilitas transisi untuk memilih kromosom yang paling unggul guna mendapatkan solusi yang optimal. (Tahir & Sugianto, 2024)

Algoritma Genetika merupakan algoritma pencarian yang didasarkan pada mekanisme seleksi alamiah dan genetika alamiah. Algoritma genetika melakukan pencarian terhadap solusi optimal berupa individu dengan kualitas terbaik dari suatu populasi. Seleksi fitur merupakan salah satu tahapan pra-proses klasifikasi. Seleksi fitur dilakukan dengan cara memilih fitur-fitur yang relevan yang mempengaruhi hasil klasifikasi. Seleksi fitur digunakan untuk mengurangi dimensi data dan fitur-fitur yang tidak relevan. Seleksi fitur digunakan untuk meningkatkan efektifitas dan efisiensi kinerja dari algoritma klasifikasi (Buani, 2021)

Algoritma yang digunakan untuk seleksi atribut adalah algoritma genetika. Metode Klasifikasi Metode Algoritma C4.5 akan diterapkan untuk klasifikasi. Selanjutnya, Algoritma C4.5 akan diperkuat menggunakan algoritma genetika untuk mengoptimalkan parameter-parameter klasifikasi. Setelah itu, kedua model akan dievaluasi menggunakan data pengujian untuk membandingkan kinerja mereka dalam mengklasifikasikan (Tahir & Sugianto, 2024).

## METODE PENELITIAN

Penelitian dilakukan dengan cara melakukan eksperimen dalam bentuk sistem penunjang keputusan untuk klasifikasi gangguan tidur memiliki tahapan sebagai berikut:

1. **Pengumpulan data.** Data yang digunakan dalam penelitian ini adalah dataset *sleep health and lifestyle* dengan jumlah data 374 record. Atribut pada dataset yaitu: *gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress level, BMI, blood pressure, heart rate, daily steps*, dan, *sleep disorder* (atribut prediktor).
2. **Pengolahan data awal.** Dilakukan proses *data validation, data integration and transformation*, dan *data size and dicritization*. Pada dataset diabetes ini semua atribut digunakan dan tidak ada data kosong.
3. **Model atau Metode yang digunakan.** Penelitian ini adalah Algoritma Decision Tree dan Algoritma Genetika sebagai seleksi fitur.
4. **Eksperimen dan Pengujian Model/Metode.** Eksperimen dilakukan dengan mengolah dataset *sleep health* menggunakan Algoritma Decision Tree dengan seleksi Algoritma Genetika. Tools yang digunakan adalah Rapidminer.
5. **Evaluasi dan Validasi Hasil.** Evaluasi dan validasi diuji tingkat akurasi dengan Confusion Matrix.

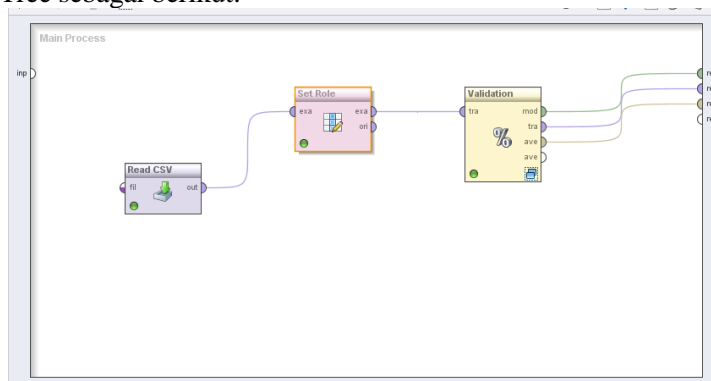
## HASIL DAN PEMBAHASAN

### Evaluasi dan Validasi

Eksperimen dan pengujian model dilakukan untuk menghitung dan mendapatkan rule-rule dari model yang diusulkan.

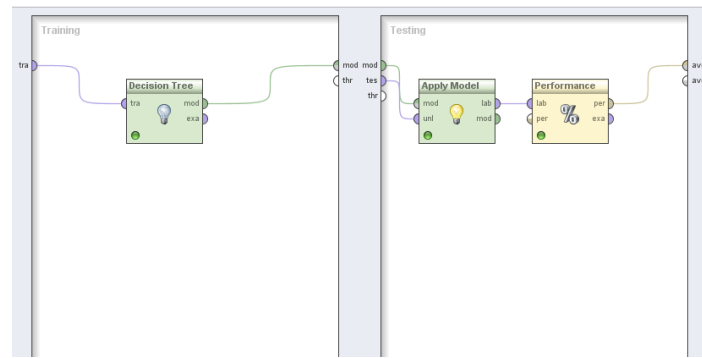
#### 1. Decision Tree

Eksperimen pertama dilakukan dengan pengujian model K-Fold Cross Validation untuk Algoritma Decision Tree sebagai berikut:



Sumber: (Septiani, 2025)

Gambar 1. Pengujian Cross Validation

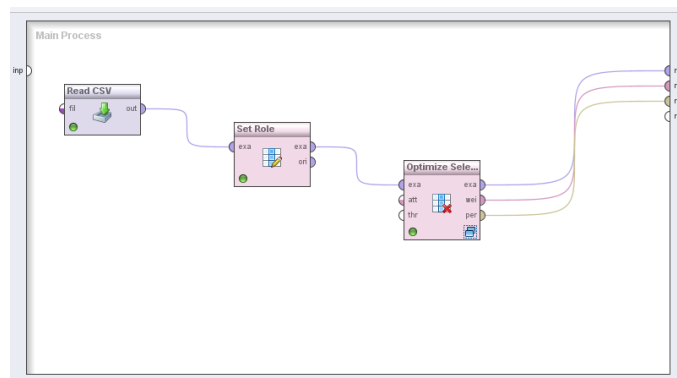


Sumber: (Septiani, 2025)

Gambar 2. Pengujian Metode Decision Tree

## 2. Algoritma Genetika

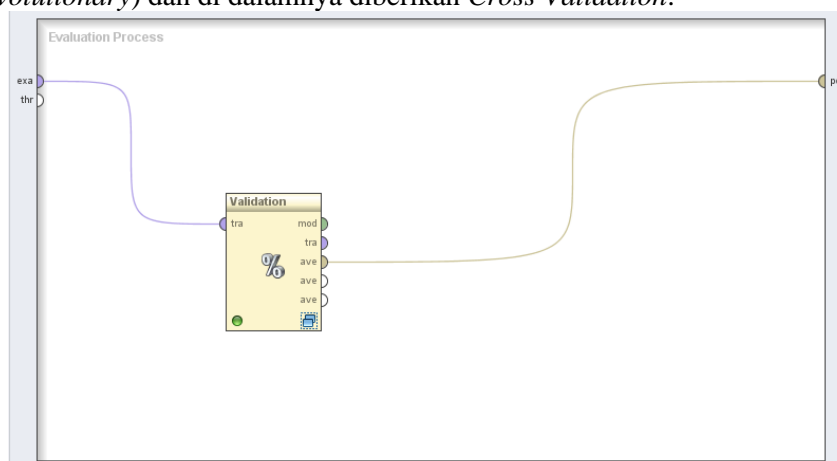
Pengujian menggunakan Decision Tree dengan fitur seleksi Algoritma Genetika sebagai berikut:



Sumber: (Septiani, 2025)

Gambar 3. *Optimize Selection (Evolutionary)*

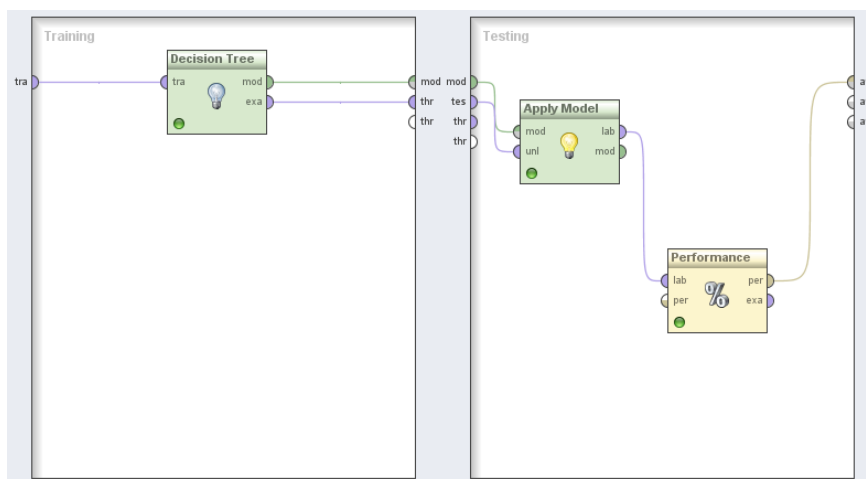
Berdasarkan gambar 3 dataset *sleep health* dihubungkan dengan *Feature Optimize Selection (Evolutionary)* dan di dalamnya diberikan *Cross Validation*.



Sumber: (Septiani, 2025)

Gambar 4. *Cross Validation*

Berdasarkan gambar 5 penggunaan *Cross Validation* dalam dataset *sleep health* terdiri dari *10-fold validation* Pada *Cross Validation* terdapat tahap dalam penggunaan algoritma *Decision Tree*.



Sumber: (Septiani, 2025)

Gambar 5. Model Decision Tree

### Hasil Penelitian

Validasi terhadap model yang diterapkan yaitu Decision Tree untuk menguji bagaimana performa dari algoritma tersebut dan menghasilkan nilai akurasi 84,82% seperti pada Tabel 1 berikut:

Tabel 1. Akurasi Algoritma Decision Tree

accuracy: 84.82%				
	true None	true Sleep Apnea	true Insomnia	class precision
pred. None	61	0	3	95.31%
pred. Sleep Apnea	5	15	3	65.22%
pred. Insomnia	4	2	19	76.00%
class recall	87.14%	88.24%	76.00%	

Sumber: (Septiani, 2025)

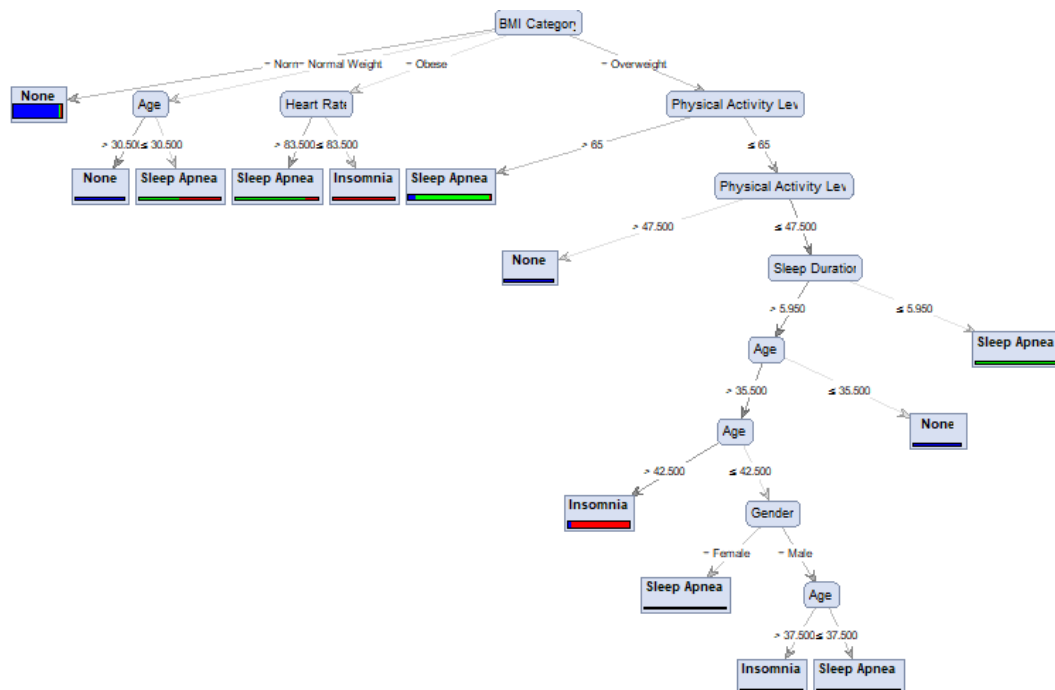
Optimasi terhadap metode Decision Tree menggunakan Algoritma Genetika menghasilkan peningkatan akurasi yaitu 95,54% seperti pada Tabel 2 berikut:

Tabel 2. Akurasi Optimasi Algoritma Genetika

accuracy: 95.54%				
	true None	true Sleep Apnea	true Insomnia	class precision
pred. None	65	0	1	98.48%
pred. Sleep Apnea	1	18	1	90.00%
pred. Insomnia	1	1	24	92.31%
class recall	97.01%	94.74%	92.31%	

Sumber: (Septiani, 2025)

Hasil dari penggunaan metode Decision Tree adalah pohon keputusan seperti pada gambar 6 sebagai berikut:



Sumber: (Septiani, 2025)

Gambar 6. Pohon Keputusan

Eksperimen ini menghasilkan 13 rule sebagai aturan dalam klasifikasi gangguan tidur. Rule yang dihasilkan adalah sebagai berikut:

1. Jika BMI = Normal Weight, maka *None* (Tidak Ada Gangguan Tidur)
2. Jika BMI = Normal Weight dan AGE > 30,5 tahun maka *None* (Tidak Ada Gangguan Tidur)
3. Jika BMI = Normal Weight dan AGE <= 30,5 tahun maka *Sleep Apnea*
4. Jika BMI = Obsese dan Heart Rate > 83,5 maka *Sleep Apnea*
5. Jika BMI = Obsese dan Heart Rate <= 83,5 maka *Insomnia*
6. Jika BMI = Overweight dan Physical Activity Level > 65 maka *Sleep Apnea*
7. Jika BMI = Overweight dan Physical Activity Level > 47,5 dan <= 65 maka *None* (Tidak Ada Gangguan Tidur)
8. Jika BMI = Overweight dan Physical Activity Level <= 47,5 dan Sleep Duration > 5,95 dan Age > 42,5 maka *Insomnia*.
9. Jika BMI = Overweight dan Physical Activity Level <= 47,5 dan Sleep Duration > 5,95 dan Age <= 42,5 dan Gender = Female, maka *Sleep Apnea*.
10. Jika BMI = Overweight dan Physical Activity Level <= 47,5 dan Sleep Duration > 5,95 dan Age > 37,5 dan Gender = Male, maka *Insomnia*.
11. Jika BMI = Overweight dan Physical Activity Level <= 47,5 dan Sleep Duration > 5,95 dan Age <= 37,5 dan Gender = Male, maka *Sleep Apnea*.
12. Jika BMI = Overweight dan Physical Activity Level <= 47,5 dan Sleep Duration <= 5,95 dan Age <= 35,5 maka *None* (Tidak Ada Gangguan Tidur).
13. Jika BMI = Overweight dan Physical Activity Level <= 47,5 dan Sleep Duration <= 5,95 maka *Sleep Apnea*.

## KESIMPULAN

*Rules* yang dihasilkan dari pengujian penggunaan fitur seleksi Algoritma Genetika yang diterapkan pada Decision Tree dapat dijadikan kontribusi dalam pengambilan keputusan terhadap klasifikasi gangguan tidur seserang dengan memperhatikan beberapa faktor yang berkaitan dengan pola tidur, gaya hidup, pola hidup sehat, juga pekerjaan dan aktivitas fisik seseorang. Dalam

pengujian menggunakan Decision Tree dengan seleksi fitur Algoritma Genetika ini menghasilkan akurasi 95,54% dan menunjukkan adanya peningkatan nilai akurasi sebesar 10,72% dari penelitian tanpa fitur seleksi yang memiliki akurasi 84,82%. Sehingga dapat disimpulkan bahwa penggunaan fitur seleksi mampu meningkatkan nilai akurasi. Penelitian ini dapat dijadikan masukan untuk dilanjutkan kembali dengan metode optimasi lain seperti Adabbost dan PSO. Berdasarkan penelitian ini dapat diberikan saran untuk diadakannya penelitian lebih lanjut dengan melakukan pengujian dengan metode lain seperti SVM, Nural Network, ataupun komparasi dari beberapa metode klasifikasi data mining.

## REFERENSI

- Buani, D. C. P. (2016). Optimasi Algoritma Naive Bayes Dengan Menggunakan Algoritma Genetika Untuk Prediksi Kesuburan (Fertility). *Jurnal Evolusi*, 4(1), 55–64.
- Buani, D. C. P. (2021). Penerapan Algoritma Naive Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung. *EVOLUSI: Jurnal Sains Dan Manajemen*, 9(2), 43–48. <https://doi.org/10.31294/evolusi.v9i2.11141>
- Gorunescu, F. (2011). *Data Mining: Concepts and Techniques*. Springer.
- Larose, D. T. (2005). *Discovering Knowledge in Database*. John Willey & Sons Inc.
- Maulida, E., Hidayati, A. F., Silalahi, F. Y., & Prasetyo, Y. D. (2024). Implementasi Algoritma K-Means Clustering dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup. *Journal of Information Engineering and Educational Technology*, 8(1), 54–61. <https://doi.org/10.26740/jieet.v8n1.p54-61>
- Nawawi, I., & Fatah, Z. (2024). Penerapan Decision Trees dalam Mendeteksi Pola Tidur Sehat Berdasarkan Kebiasaan Gaya Hidup. 2(4), 34–41.
- 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>
- Septiani, W. D. (2025). *Laporan Akhir Penelitian Mandiri*.
- Septiani, W. D., & Marlina. (2021). Comparison of Decision Tree, Naive Bayes, and Neural Network Algorithm for Early Detection of Diabetes. *Jurnal Pilar Nusa Mandiri*, 17(1), 73–78. <https://doi.org/10.33480/pilar.v17i1.2213>
- Sukri, M. H., & Handrianto, Y. (2024). Penerapan Algoritma C4. 5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta. *Informatics and Computer Engineering Journal*, 4(1), 11–24.
- Sunge, A. S. (2018). Optimasi Algoritma C4.5 Menggunakan Genetic Algoritma Dalam Memprediksi Website Phishing. *Seminar Nasional Inovasi Dan Tren (SNIT)*, 92.
- Tahir, S. F., & Sugianto, C. A. (2024). Optimasi Naive Bayes Menggunakan Algoritma Genetika Pada Klasifikasi Komentar Cyberbullying Pada Media Sosial X. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3), 3350–3356. <https://doi.org/10.23960/jitet.v12i3.4834>