

- 1. Dalam algoritma Decision Tree, komponen manakah yang mewakili seluruh populasi atau sampel data sebelum dilakukan pemecahan pertama?**

  - A. Root Node (Akar)
  - B. Internal Node (Cabang)
  - C. Splitting
  - D. Leaf Node (Daun)
- 2. Apa yang diindikasikan oleh nilai Entropy yang tinggi (mendekati 1) pada sebuah node dalam Decision Tree?**

  - A. Model mengalami underfitting pada node tersebut.
  - B. Node tersebut sangat murni dan hanya berisi satu kelas data.
  - C. Node tersebut merupakan node akhir (leaf node) yang optimal.
  - D. Node tersebut sangat tercampur dan tidak murni.
- 3. Dalam studi kasus yang dibahas, mengapa dataset dibagi menjadi Train Set (80%) dan Test Set (20%)?**

  - A. Agar model dapat belajar dari dua jenis data yang berbeda secara bersamaan.
  - B. Untuk mengurangi jumlah fitur yang perlu dianalisis oleh model.
  - C. Untuk mempercepat proses training dengan menggunakan data yang lebih sedikit.
  - D. Untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya.
- 4. Hyperparameter `max\_depth` digunakan untuk mengontrol kedalaman pohon. Apa risiko utama jika `max\_depth` diatur dengan nilai yang sangat besar?**

  - A. Underfitting, di mana model terlalu sederhana untuk menangkap pola data.
  - B. Waktu prediksi yang lebih lambat pada data baru.
  - C. Ketidakmampuan model untuk menggunakan Gini Index.
  - D. Overfitting, di mana model terlalu detail dan 'menghafal' data latihan.
- 5. Metode mana yang digunakan untuk mengukur kualitas pemisahan (split) dan menjadi pilihan `criterion` pada studi kasus tersebut?**

A. Entropy

B. Gini

C. Gain Ratio

D. Information Gain

**6. Menurut dokumen, tujuan utama dari Cost Function dalam Decision Tree adalah untuk meminimalkan apa?**

A. Waktu komputasi

B. Akurasi model

C. Impurity (ketidakmurnian)

D. Jumlah node dalam pohon

**7. Jika sebuah node memiliki nilai Gini Impurity sebesar 0, apa artinya?**

A. Node tersebut berisi data dari semua kelas dengan proporsi yang sama.

B. Model tersebut memiliki akurasi 0%.

C. Node tersebut tidak dapat digunakan untuk membuat prediksi.

D. Node tersebut murni, artinya semua sampel di dalamnya berasal dari satu kelas saja.

**8. Apa fungsi dari Gain Ratio dalam algoritma C4.5?**

A. Mengatasi kelemahan Information Gain yang bias terhadap fitur dengan banyak nilai unik.

B. Untuk mempercepat perhitungan dengan menyederhanakan rumus Entropy.

C. Untuk memastikan nilai Entropy selalu positif.

D. Untuk menggantikan Gini Index dalam algoritma CART.

**9. Berdasarkan bagian 'Dasar Penyusunan Hipotesis', hipotesis alternatif () didukung oleh asumsi bahwa...**

A. Dataset Breast Cancer memiliki fitur-fitur yang relevan secara klinis.

B. Data tidak perlu dibagi menjadi set latihan dan pengujian.

C. Hyperparameter tidak diperlukan untuk mencapai akurasi tinggi.

D. Decision Tree adalah algoritma yang paling cepat untuk data medis.

**10. Dalam studi kasus, variabel target (Y) untuk diagnosis 'Ganas (Malignant)' diberi label sebagai...**

- A. 0
- B. Malignant
- C. 1
- D. -1