

Nama : Ilham Novriadi
NIM : 231011403539

Bagian 1 – Pemahaman Konsep (Teori)

1. Apa yang dimaksud dengan Decision Tree?

Decision Tree adalah algoritma *supervised learning* non-parametrik yang digunakan untuk tugas klasifikasi dan regresi.

Algoritma ini memodelkan keputusan dan konsekuensi yang mungkin terjadi dalam struktur seperti pohon (*tree-like structure*), di mana:

- Setiap **simpul internal** mewakili pengujian pada atribut.
- Setiap **cabang** mewakili hasil pengujian.
- Setiap **simpul daun** (*leaf node*) memegang label kelas atau nilai target.

2. Penjelasan Konsep Penting

Berikut adalah istilah-istilah kunci dalam struktur Decision Tree:

- Node
Titik keputusan dalam pohon di mana data dipisahkan berdasarkan fitur tertentu.
- Root (Akar)
Node paling atas dari pohon (awal mula) yang berisi seluruh dataset sebelum percabangan apa pun.
- Leaf (Daun)
Node terminal atau ujung dari pohon yang tidak memiliki cabang lagi. Ini merepresentasikan hasil prediksi atau keputusan akhir.
- Splitting
Proses membagi sebuah node menjadi dua atau lebih sub-node berdasarkan kriteria tertentu (misalnya Gini Impurity atau Entropy) untuk memisahkan kelas target sebaik mungkin.
- Pruning
Proses pemangkasan cabang pohon yang tidak signifikan atau terlalu kompleks untuk mencegah overfitting.
 - *Pre-pruning*: Dilakukan saat pembentukan pohon.
 - *Post-pruning*: Dilakukan setelah pohon terbentuk.

3. Perbedaan Decision Tree, Random Forest, dan Gradient Boosting

Tabel berikut merangkum perbedaan utama antara ketiga metode ini:

Fitur	Decision Tree	Random Forest	Gradient Boosting
Definisi	Satu pohon keputusan tunggal.	Kumpulan banyak pohon keputusan (<i>Bagging</i>).	Kumpulan pohon yang dibangun secara berurutan (<i>Boosting</i>).
Cara Kerja	Membagi data secara rekursif hingga kriteria berhenti terpenuhi.	Membangun banyak pohon secara independen (paralel) dan mengambil suara mayoritas (voting).	Membangun pohon secara sekuensial, di mana pohon baru memperbaiki kesalahan pohon sebelumnya.
Kecenderungan	Cenderung <i>overfitting</i> (varian tinggi).	Mengurangi varian (lebih stabil & akurat).	Mengurangi bias dan varian (sangat akurat).
Kecepatan	Cepat dilatih & diprediksi.	Lambat dilatih (karena banyak pohon), namun cepat diprediksi.	Lambat dilatih (sekuensial), cukup cepat diprediksi.

4. Kelebihan dan Kekurangan Tree-Based Methods

Kelebihan

- Mudah Diinterpretasikan:** Struktur pohon (khususnya Decision Tree tunggal) mudah dipahami dan divisualisasikan oleh manusia.
- Tidak Butuh Penskalaan:** Tidak memerlukan *Normalization* atau *Standardization* pada data.
- Fleksibel:** Dapat menangani data numerik dan kategorikal sekaligus.
- Tahan Outlier:** Lebih tahan terhadap outlier dibandingkan model linear.

Kekurangan

- Overfitting:** Decision Tree tunggal sangat rentan membuat model yang terlalu kompleks (*overfitting*).
- Ketidakstabilan:** Perubahan kecil pada data latih bisa menghasilkan struktur pohon yang sangat berbeda.
- Bias pada Kelas Tidak Seimbang:** Cenderung bias jika dataset memiliki *imbalanced class* (satu kelas mendominasi).