

Decision tree

Decision tree adalah model prediktif dalam analisis data yang menggunakan struktur pohon keputusan untuk memvisualisasikan dan membuat keputusan berdasarkan serangkaian aturan. Ini adalah algoritma pembelajaran mesin yang populer karena kemampuannya dalam memahami dan memodelkan hubungan kompleks antara fitur-fitur data.

Classification tree sendiri adalah jenis decision tree yang khusus digunakan untuk masalah klasifikasi, di mana tujuannya adalah untuk mengklasifikasikan data ke dalam beberapa kelas atau kategori yang telah ditentukan sebelumnya. Dengan menggunakan classification tree, model dapat mempelajari pola-pola dalam data untuk membuat prediksi tentang kelas mana yang paling mungkin untuk suatu sampel data tertentu.

Terminologi :

1. **Simpul** : Representasi titik dalam pohon keputusan dimana keputusan dibuat atau hasil diprediksi. Ada dua jenis simpul: simpul internal dan simpul daun.
2. **Simpul Internal** : Simpul dimana keputusan dibuat berdasarkan pada fitur-fitur atau atribut-atribut tertentu.
3. **Simpul Daun** : Simpul yang tidak memiliki cabang lagi dalam pohon. Ini mewakili hasil atau prediksi akhir dari model.
4. **Cabang** : Garis atau sambungan yang menghubungkan satu simpul ke simpul lainnya dalam pohon keputusan.
5. **Atribut** : Fitur-fitur atau variabel-variabel yang digunakan untuk membuat keputusan di setiap simpul.

Gini impurity adalah salah satu metrik yang digunakan dalam pembentukan pohon keputusan dalam analisis data dan pembelajaran mesin. Ini mengukur seberapa sering elemen dalam satu kelas tertentu akan salah diidentifikasi jika dipilih secara acak. Dalam konteks pohon keputusan, Gini impurity digunakan untuk mengevaluasi bagaimana sebuah node memisahkan kelas target ke dalam subgrup yang berbeda.