

Nama: Irawan Mardiansyah

Kelas: TK-45-04

NIM: 1103210174

Decision Tree merupakan sebuah rancangan yang digunakan untuk membuat keputusan berdasarkan serangkaian pertanyaan dan juga jawaban. Pertanyaan-pertanyaan ini disusun dengan bentuk cabang-cabang pohon, dan jawabannya menentukan arah mana yang akan diambil berikutnya. Pada umumnya Decision Tree membuat pernyataan dan kemudian membuat keputusan berdasarkan benar atau tidaknya pernyataan itu, masalah ketika Decision Tree mengklasifikasikan hal-hal ke dalam kategori itu disebut pohon klasifikasi dan ketika Decision Tree memprediksi nilai numerik itu disebut pohon regresi. Pohon klasifikasi itu menggabungkan data numerik, dengan tidak ada data jadi tidak masalah untuk mencampur tipe data di pohon yang sama. Untuk sebagian besar pohon klasifikasi cukup mudah untuk dikerjakan, mulai dari atas sampai bawah. Dalam Decision Tree terdapat tiga struktur yang utama, yaitu:

- Node: titik keputusan dalam pohon, yang berfungsi sebagai titik percabangan dari suatu proses pengambilan keputusan yang ada.
- Branches (cabang): mewakili berbagai hasil yang berasal dari sebuah node. Setiap cabang dihubungkan dengan kondisi tertentu lainnya.
- Leaves (daun): mewakili hasil akhir berdasarkan serangkaian keputusan yang dibuat sebelumnya. Leaves biasanya berisi klasifikasi kelas atau prediksi.

Gini Impurity merupakan metrik yang digunakan untuk mengukur suatu ketidakmurnian dari node dalam decision tree. Semakin tinggi gini impurity, semakin tidak murni node tersebut. gini impurity membantu algoritma decision tree untuk membagi data secara optimal. Algoritma ini akan memilih atribut yang menghasilkan gini impurity paling tinggi. Hal tersebut menghasilkan decision tree yang lebih efektif dan akurat dalam memprediksi suatu hasil. Algoritma decision tree akan memilih atribut yang paling tinggi untuk membagi data node. Atribut dengan gini impurity paling tinggi yaitu atribut paling informatif dalam membagi data menjadi bagian yang lebih beragam.

Memilih variabel tepat untuk decision tree bisa menggunakan dua jenis variabel: numerik dan juga kontinu. Variabel numerik biasanya terdapat nilai angka, seperti penilaian. Sedangkan variabel kontinu memiliki nilai pecahan dan tidak terbatas pada rentang tertentu, contohnya suhu. Biasanya variabel kontinu diubah menjadi variabel klasifikasi terlebih dahulu. Hal ini dilakukan dengan membagi variabel kontinu menjadi beberapa interval. Variabel numerik dan kontinu digunakan dalam decision tree, hal ini bisa membuat percabangan yang membantu dalam proses klasifikasi. Pemilihan variabel yang tepat dan pembagian yang ideal sangat penting agar bisa menghasilkan model decision tree yang akurat.