

Nama: Nella Aprilia
NIM: 1103210185
Kelas: TK-45-04
Decision and Classification Trees

Decision tree secara umum adalah untuk membuat pernyataan dan kemudian membuat keputusan berdasarkan apakah pernyataan itu benar atau tidak, itu bukan masalah besar ketika decision tree mengklasifikasikan sesuatu ke dalam kategori yang disebut pohon klasifikasi dan kapan decision tree memprediksi nilai numerik yang disebut pohon regresi. Pohon klasifikasi menggabungkan data numerik dengan mencampur tipe data jika tidak ada data tidak menjadi masalah. Struktur decision tree terdapat 3 struktur yaitu, Node merupakan simpul akar atau hanya akarnya saja yang disebut, Cabang Memiliki panah yang menunjuk dan mengarahkan, Daun Memiliki panah yang menunjuk tetapi tidak ada anak panah yang mengarah ke luar.

Gini Impurity adalah sebuah metrik yang berperan penting dalam algoritma decision tree. Fungsinya adalah untuk mengukur tingkat ketidaksempurnaan data pada setiap node dalam tree. Semakin tinggi nilai gini impurity, semakin tidak sempurna data pada node tersebut. Nilai gini impurity dapat membantu algoritma decision tree dalam menentukan cara pembagian data yang optimal. Algoritma ini akan memilih atribut yang menghasilkan nilai gini impurity paling tinggi. Hal ini akan menghasilkan decision tree yang lebih efektif dan akurat dalam memprediksi suatu hasil. Dengan menggunakan gini impurity, Decision tree dapat menghasilkan model yang lebih optimal dan akurat dalam memprediksi berbagai jenis data.

Membangun decision tree yang efektif memerlukan pemilihan variabel yang tepat. Dua jenis variabel yang umum digunakan adalah numerik dan kontinu. Variabel numerik yang memiliki penilaian sedangkan variabel kontinu memiliki nilai pecahan. Pada umumnya, variabel kontinu diubah menjadi variabel klasifikasi terlebih dahulu. Hal ini dilakukan dengan membagi variabel kontinu menjadi beberapa interval. Manfaat Penggunaan variabel numerik dan kontinu pada penggabungan variabel numerik dan kontinu dalam decision tree yaitu dapat menghasilkan percabangan yang lebih optimal dan membantu proses klasifikasi. Pentingnya Pemilihan dan pembagian pada variabel untuk memilih variabel yang tepat dan menentukan pembagian interval yang ideal merupakan langkah penting untuk menghasilkan model decision tree yang terpercaya.