

Dheandra Alfarrelwijaya

1103213048

TK-45-04

### Decision Tree StatQuest

Decision Tree adalah algoritma yang membuat pernyataan dan mengambil keputusan berdasarkan kebenaran pernyataan tersebut. Ketika digunakan untuk klasifikasi, algoritma ini membentuk pohon klasifikasi, dan untuk prediksi nilai numerik, algoritma ini membentuk pohon regresi. Proses pembentukan pohon dimulai dari simpul akar dan turun ke simpul daun. Untuk membangun pohon klasifikasi, pertanyaan di bagian atas pohon dipilih berdasarkan kemampuan variabel untuk memprediksi kategori tertentu.

Di dalam video menjelaskan konsep 'gini impurity' untuk mengukur ketidakmurnian prediksi pada Decision Tree. Ini diilustrasikan dengan contoh prediksi preferensi film, di mana gini impurity dihitung untuk setiap daun yang mencakup campuran preferensi. Metode lain seperti entropi dan perolehan informasi juga disebutkan tetapi fokus pada gini impurity karena kepopulerannya. juga menjelaskan bagaimana menghitung gini impurity dengan mengurutkan nilai umur dan menghitung rata-rata umur untuk menentukan impurity untuk setiap kelompok umur. Selanjutnya, dijelaskan bagaimana nilai gini impurity dihitung untuk setiap fitur guna menentukan urutan fitur dalam pohon klasifikasi.

Overfitting dalam Decision Tree juga disertakan, dengan diperkenalkannya dua metode untuk mengatasinya: pemangkasan dan pembatasan pertumbuhan pohon. Validasi silang juga disebutkan sebagai cara untuk menguji efektivitas berbagai metode dalam menangani overfitting.