

Nama : Ilham Khodar Trijaya  
NIM : 1103213084  
Kelas : TK-45-04

Umumnya, pohon keputusan membuat pernyataan dan mengambil keputusan berdasarkan apakah pernyataan tersebut TRUE (benar) atau FALSE (salah). Jika pohon keputusan membagi sesuatu menjadi beberapa kategori, maka disebut pohon klasifikasi, dan jika pohon keputusan memprediksi angka, maka disebut pohon regresi.

Bagian-bagian yang menyusun Decision Tree :

1. Simpul Akar (Root Node): Ini adalah titik awal pohon dan mewakili keputusan atau pertanyaan utama yang perlu dijawab. Biasanya digambarkan sebagai simpul paling atas di pohon.
2. Cabang (Branch): Garis-garis ini menghubungkan simpul dan mewakili berbagai kemungkinan jawaban atau hasil dari pertanyaan di setiap simpul.
3. Simpul Daun (Leaf Node): Ini adalah simpul terminal pohon dan mewakili keputusan akhir atau klasifikasi untuk kumpulan data tertentu. Biasanya digambarkan sebagai kotak atau lingkaran di ujung cabang.

Membuat Decision Tree menggunakan data mentah, terdapat beberapa klasifikasi yang akan diprediksi antara lain 'Loves Popcorn', 'Loves Soda', 'Age'. Setelah didapatkan Decision Tree sementara, banyaknya orang dalam data hampir menyukai dan tidak menyukai 3 klasifikasi diatas, sehingga dapat dilakukan metode Gini Impurity.

Berdasarkan 3 klasifikasi Tersebut maka akan:

1. Menghitung Gini Impurity untuk Loves Popcorn
2. Menghitung Gini Impurity untuk Loves Soda
3. Menghitung Gini Impurity untuk Age

Gini impurity adalah metrik yang digunakan untuk mengukur tingkat ketidak homogenan (keseragaman) di dalam simpul (node). Semakin rendah Gini impurity, semakin homogen simpul tersebut, artinya data pada simpul tersebut semakin condong ke satu kelas tertentu. Dengan kata lain, Gini impurity membantu menentukan pembelahan data yang optimal di setiap simpul pohon keputusan. Pembelahan yang baik akan menghasilkan simpul anak (child node) yang lebih homogen dibandingkan dengan simpul induknya (parent node).

Metode pembuatan pohon klasifikasi berpusat pada penurunan ketidakmurnian di setiap langkah. Biasanya dilakukan dengan memilih langkah yang menghasilkan Gini Impurity paling rendah, seperti mengisolasi informasi berdasarkan kecenderungan terhadap 'Loves Popcorn', 'Loves Soda', 'Age'. Pentingnya mengantisipasi overfitting juga ditekankan untuk menjamin ekspektasi yang tepat.

Untuk mengatasi overfitting, digunakan prosedur pemangkasan dan pemotongan perkembangan pohon. Ini termasuk membuang cabang-cabang yang tidak penting dari pohon atau membatasi jumlah informasi di setiap daun. Langkah-langkah ini menawarkan bantuan untuk membuat model yang lebih umum dan dapat dihubungkan dengan baik ke informasi modern dalam pembelajaran mesin.