Konsep Dasar Decision Tree

Konsep dari decision tree adalah mengubah data menjadi aturan-aturan keputusan. Manfaat utama dari penggunaan decision tree adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Selain itu, decision tree juga dapat menangani data yang tidak lengkap atau mengandung nilai yang hilang. Namun, kelemahannya adalah kecenderungan untuk overfitting jika pohon terlalu kompleks atau jika dataset memiliki banyak fitur yang tidak relevan.

Decision tree juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Decision tree memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan. Decision tree digunakan dalam berbagai bidang, termasuk ilmu komputer, statistik, dan pembelajaran mesin.

Nama lain dari decision tree adalah CART (Classification and Regression Tree). Dimana metode ini merupakan gabungan dari dua jenis pohon, yaitu classification tree dan juga regression tree. Jika variabel dependen yang dimiliki bertipe kategorik maka CART menghasilkan pohon klasifikasi (classification trees). Sedangkan jika variabel dependen yang dimiliki bertipe kontinu atau numerik maka CART menghasilkan pohon regresi (regression trees).

Decision tree ini bisa terjadi overlap, terutama ketika kelas dan kriteria yang digunakan sangat banyak tentu saja dapat meningkatkan waktu pengambilan keputusan sesuai dengan jumlah memori yang dibutuhkan.

Dalam hal akumulasi, decision tree juga seringkali mengalami kendala eror terutama dalam jumlah besar. Selain itu, terdapat pula kesulitan dalam mendesain decision tree yang optimal. Apalagi mengingat kualitas keputusan yang didapatkan dari metode decision tree sangat tergantung pada bagaimana pohon tersebut didesain. Maka dibutuhkan Random Forest untuk mengatasi overlap di atas.

Demikianlah konsep dasar dari decision tree. Metode ini memiliki banyak variasi dan pengembangan lebih lanjut, seperti decision tree ensemble (seperti Random Forest) dan decision tree dengan variabel target kontinu (seperti Regression Tree).

Berikut adalah langkah-langkah dasar dalam membangun dan menggunakan decision tree:

1. Pembentukan Node Awal

Decision tree dimulai dengan node awal, juga dikenal sebagai root node. Node ini mewakili seluruh dataset yang tersedia.

1. Pemilihan Fitur

Pada setiap langkah dalam decision tree, kita memilih fitur yang paling relevan untuk membagi dataset. Pemilihan fitur ini didasarkan pada beberapa metode, seperti indeks keuntungan (misalnya, Information Gain), indeks Gini, atau chi-square.

1. Pemisahan Dataset

Setelah fitur dipilih, dataset dibagi berdasarkan nilai fitur tersebut. Setiap cabang dari node menggambarkan kemungkinan nilai fitur yang berbeda.

1. Pengulangan Langkah 2 dan 3

Proses pemilihan fitur dan pemisahan dataset diulang secara rekursif untuk setiap cabang baru yang dihasilkan. Hal ini dilakukan hingga dataset pada suatu cabang telah sepenuhnya terklasifikasi atau tidak dapat dibagi lebih lanjut berdasarkan fitur yang tersedia.

1. Penentuan Label Kelas

Setelah semua pemisahan dilakukan, label kelas ditentukan untuk setiap cabang akhir atau daun (leaf) berdasarkan mayoritas sampel yang ada di cabang tersebut.

1. Pruning (Pemangkasan) Pohon

Proses pruning adalah langkah opsional yang melibatkan pengurangan ukuran pohon keputusan dengan menghapus cabang-cabang yang tidak signifikan. Hal ini dapat membantu mencegah overfitting dan meningkatkan generalisasi model.

1. Pengambilan Keputusan

Setelah decision tree dibangun, kita dapat menggunakan pohon tersebut untuk mengambil keputusan dengan memasukkan data baru ke dalam pohon dan mengikuti jalur yang sesuai dengan nilai fitur.