

Nama : M Rayyan Aqilah Manna

NIM : 1103210210

Decision and Classification Tree

Decision Tree (pohon keputusan) adalah model prediktif dalam machine learning yang digunakan untuk mengambil keputusan berdasarkan serangkaian aturan logis yang dibentuk dalam struktur pohon. Pohon keputusan memodelkan keputusan dan konsekuensi secara hierarkis, di mana setiap simpul pada pohon mewakili suatu keputusan atau pengujian terhadap atribut tertentu, dan setiap cabang dari simpul menggambarkan hasil dari pengujian tersebut.

Saat decision tree mengklasifikasikan sesuatu menjadi kategori, itu termasuk classification tree. Tetapi jika decision tree memprediksi numeric value, maka itu termasuk regression tree. Classification tree yang kompleks juga dapat menggabungkan data numerik dengan “Yes” dan “No” data. Tree ini dapat menanyakan sesuatu yang sama dengan perbedaan jumlah sehingga batas numeric bisa saja berbeda pada data yang sama, dan klasifikasi final dapat diulang.

Kita dapat mengukur impurity dari hasil prediksi sebuah tree dengan menggunakan gini impurity. Gini impurity bukan satu satunya cara untuk menghitung impurity, terdapat cara lain seperti entropi tetapi gini impurity lebih populer dan juga lebih mudah untuk dilakukan. Menghitung gini impurity dimulai dengan menghitung impurity setiap daun, lalu menghitung total rata-rata gini impurity dari semua daun.

Untuk menghitung gini impurity dari sebuah data numerik akan sedikit berbeda dari sebelumnya, kita akan memulai dengan mengurutkan dari kecil ke besar, lalu menghitung rata-rata dari sebuah data dan data setelahnya. Dari data rata-rata tersebut, kita akan melakukan perhitungan sama seperti bentuk data sebelumnya untuk mencari gini impurity terkecil. Dari gini impurity terkecil tersebut, kita dapat membandingkannya dengan gini impurity lain dan kategori apapun yang memiliki impurity terkecil akan dijadikan puncak tree yang dibuat. Jika daun dari puncak pohon tersebut masih memiliki impurity, kita dapat memisahkannya lagi menjadi daun baru sampai tidak memiliki impurity.