

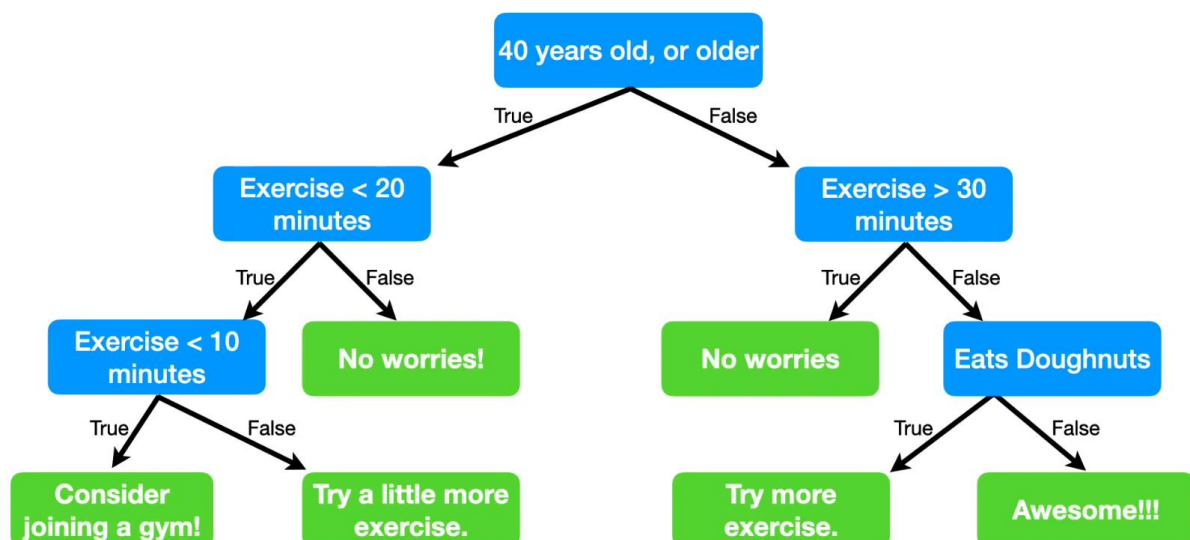
Nama : Fadhil Dzikri Aqila

NIM : 1103213136

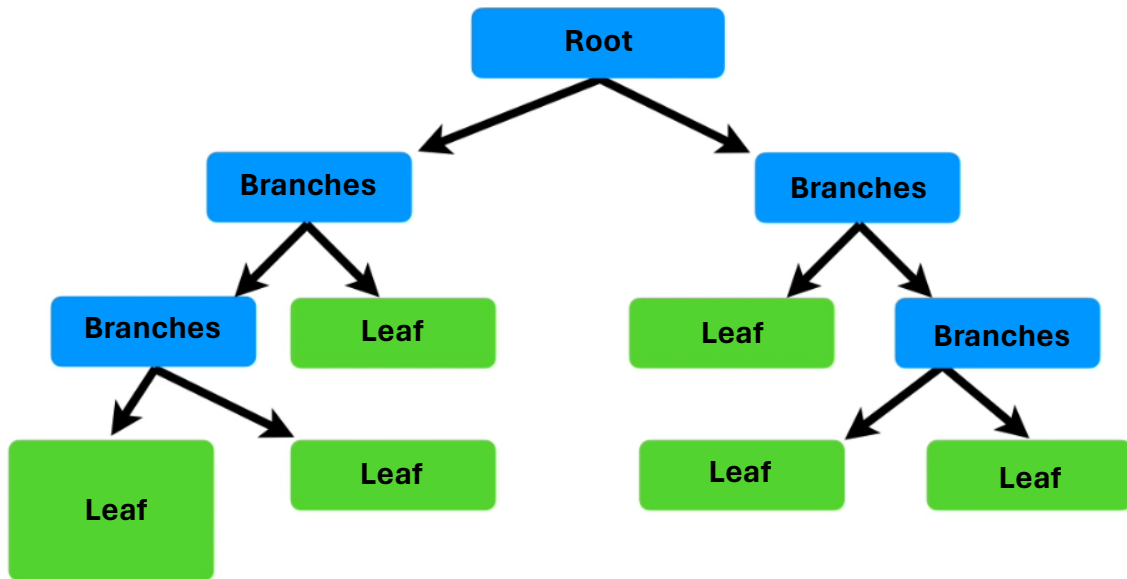
Kelas : TK-45-03

## Machine Learning – Decision and Classification Tree

Secara umum decision tree membuat suatu pernyataan dan kemudian mengambil keputusan berdasarkan benar atau salahnya pernyataan tersebut. Ketika decision tree mengklasifikasikan berbagai hal ke dalam kategori, hal itu disebut sebagai classification tree. Dan ketika decision tree memprediksi nilai numerik, hal itu disebut regression tree.



Gambar diatas adalah contoh classification tree rumit yang menggabungkan data numerik dan data true/false. Jadi tidak apa-apa untuk menggabungkan tipe data dalam tree yang sama. Tree tersebut bertanya tentang exercise beberapa kali dan jumlah waktu exercise nya tidak selalu sama, jadi ambang batas numerik bisa berbeda untuk data yang sama. Hasil akhir klasifikasi juga dapat diulang misalnya terdapat 2 “No worries!” pada tree diatas.



Gambar diatas adalah contoh struktur decision tree. Bagian paling atas dari tree disebut root atau akar. Cabang dari root disebut branches atau cabang, cabang memiliki panah masuk dan panah keluar. Terakhir ada Leaf atau daun, leaf mempunyai panah masuk tetapi tidak ada panah keluar.

Ada beberapa cara untuk menghitung Impurity dari leaf, salah satu metode yang paling populer disebut gini impurity. Selain metode tersebut ada juga metode lain yaitu entropy dan information gain. Gini impurity adalah ukuran yang digunakan dalam pengklasifikasian untuk mengukur seberapa seragam (homogen) sebuah himpunan data dari segi label kelasnya. Gini impurity untuk sebuah himpunan data dihitung dengan rumus :

$$Gini(t) = 1 - \sum_{i=1}^j P(i|t)^2$$

Gini impurity berkisar dari 0 sampai 0,5. Nilai 0 menunjukkan bahwa himpunan data sepenuhnya sama atau seragam, sedangkan nilai 0,5 menunjukkan bahwa himpunan data bermacam-macam dan tidak seragam.

Ketika membangun decision tree, Gini Impurity digunakan sebagai kriteria pemiliha node untuk membagi himpunan data. Tujuannya adalah untuk meminimalkan Gini impurity setelah pembagian, sehingga menghasilkan pemisahan yang lebih baik antara kelas-kelas yang ada.