

# CART Decision Tree

## Procedure 方法

CART Decision Tree adalah suatu model machine learning berbasis tree, yang bekerja dengan cara memecah data sambil mendapatkan information gain dan mengurangi impurity.

1. **Inisialisasi**

Model di-inisialisasi dengan membuat root node yang berisi data yang ingin di-fit.

2. **Mencari best question**

Question adalah sebuah conditional berdasarkan satu fitur, yang akan membelah data menjadi dua. Best question ini dicari secara brute force, dan question yang membagi data dengan information gain terbesar (gini impurity sekarang dikurangi dengan weighted gini impurity pembagian kiri kanan), dipilih menjadi best question.

3. **Buat Node kiri kanan dari question**

Node kiri dan kanan dibuat dengan data berdasarkan conditional dari question.

4. **Ulangi hingga max depth atau node tidak bisa di-split**

5. **Prediksi**

Prediksi dilakukan dengan menelusuri pohon sesuai dengan conditionalnya, dan jika sampai leaf node, dilakukan voting untuk classification dan mean untuk regression.

## VS Sklearn

Hasil di notebook DoE menghasilkan nilai F1 yang sedikit lebih rendah. Hal tersebut mungkin dikarenakan sklearn memilih splitter question yang lebih pintar atau pruning.

## Potential Improvements

- Pruning agar pohon tidak overfit
- Splitting criterion yang lain selain gini impurity