

Decision Tree Classifier:

1. Pertama, dataset digabungkan menjadi satu matriks. Kemudian, akan dibangun pohon secara rekursif dengan membagi dataset menjadi dua subset berdasarkan fitur dan *threshold* yang memberikan information gain terbaik. Pembelahan pohon akan berhenti jika salah satu kondisi ini terpenuhi:

- Jumlah sampel di suatu node $< \text{min_sample_split}$
- Kedalaman maksimum sudah tercapai (max_depth)
- Tidak ada lagi informasi gain yang signifikan dari hasil *splitting* lebih lanjut

Setelah itu, akan dipilih *split* yang terbaik. *Split* terbaik adalah *split* yang memaksimalkan information gain, yang diukur dengan perbedaan impurity sebelum dan sesudah split. Setelah model selesai dibangun, prediksi dari data baru dilakukan dengan menjalankan data tersebut ke dalam pohon mengikuti jalur dari *root node* ke *leaf* yang telah dibuat sebelumnya berdasarkan fitur-fitur dari data yang diberikan.

2. Model Decision Tree Classifier dari scikit-learn memiliki akurasi yang lebih tinggi. Hal ini karena Decision Tree Classifier scikit learn dirancang untuk meminimalkan kesalahan perhitungan yang muncul karena presisi floating-point. Ini penting dalam perhitungan informasi gain atau dalam pemilihan split, terutama ketika dataset memiliki banyak fitur atau distribusi nilai yang luas. Selain itu, *pruning tree* pada scikit learn juga dilakukan lebih baik dibandingkan model dari *scratch* yang telah dibikin. Terakhir, model

Decision Tree dari *scratch* juga kurang optimal, hal ini terlihat pada waktu eksekusi yang jauh lebih lama dibandingkan model dari scikit learn.

3. Hal yang bisa ditingkatkan dari model Decision tree *scratch* saat ini adalah optimisasi algoritma dan menambahkan *hyperparameter tuning* lainnya sehingga waktu eksekusi yang dibutuhkan serta tingkat akurasi yang dihasilkan bisa mendekati model Decision Tree scikit learn yang asli.