

Decision Trees

Bab 6 membahas konsep dasar Decision Tree sebagai model yang bekerja dengan membelah ruang fitur menjadi area-area yang lebih kecil berdasarkan serangkaian aturan sederhana. Algoritma ini memilih fitur dan *threshold* terbaik untuk pembelahan menggunakan kriteria *impurity* seperti Gini impurity atau entropy. Proses pembelahan yang berulang ini membentuk struktur pohon dengan *node* keputusan dan *leaf node* yang memuat prediksi akhir

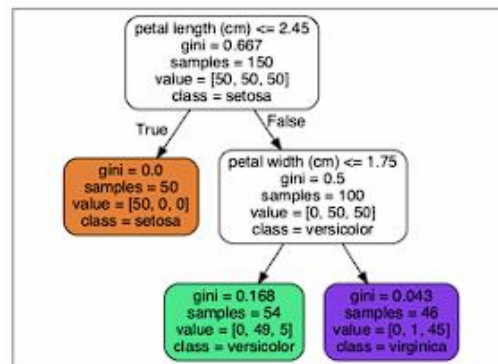


Figure 6-1. Iris Decision Tree

Kekuatan *Decision Tree* terletak pada kemampuannya untuk menangani pola non-linier, namun model ini sangat rentan terhadap overfitting jika pembelahan dibiarkan berlanjut hingga kedalaman maksimal, menghasilkan batas keputusan yang terlalu ketat mengikuti data *training*.

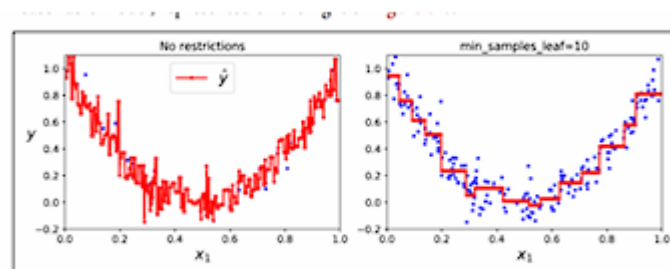


Figure 6-6. Regularizing a Decision Tree regressor

Untuk mengatasi *overfitting*, digunakan teknik pruning (pemangkasan) atau regularisasi, seperti membatasi kedalaman pohon (*max_depth*), atau membatasi jumlah sampel minimum yang diperlukan dalam sebuah *leaf* (*min_samples_leaf*). Selain klasifikasi, *Decision Tree* juga dapat digunakan untuk regresi (*Decision Tree Regressor*), di mana pohon membagi ruang data dan menghasilkan nilai rata-rata target pada setiap *leaf*. Meskipun kuat, *Decision Tree* memiliki keterbatasan, yaitu sensitivitas terhadap rotasi data dan instabilitas terhadap sedikit perubahan pada *dataset*. Untuk mengatasi kekurangan ini, bab selanjutnya akan memperkenalkan metode ensemble.