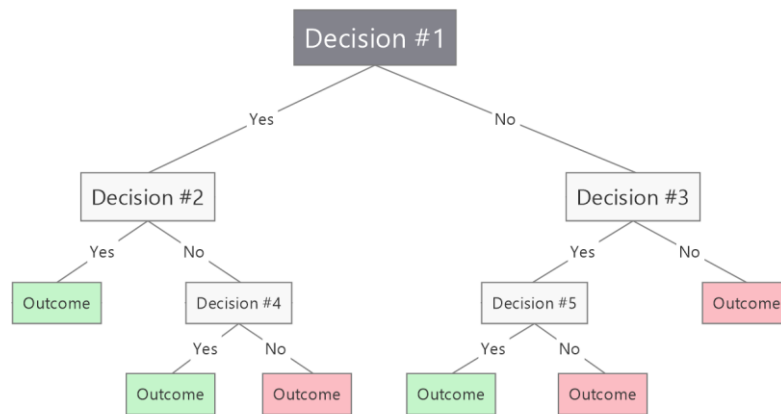


Mohamad Rayi Dwi Putra

1103213070

TK45-04

Basic Decision Tree Concepts



Pohon keputusan adalah alat pembelajaran mesin yang termasuk dalam kategori supervised learning. Ia digunakan untuk membuat prediksi berdasarkan serangkaian aturan keputusan. Pohon keputusan ini disusun seperti pohon, dengan:

- Akar (Root): mewakili pertanyaan awal yang diajukan untuk memulai proses pengambilan keputusan.
- Cabang (Branches): mewakili jawaban yang mungkin untuk pertanyaan tersebut. Setiap cabang mengarah ke simpul anak (child node).
- Simpul Anak (Child Node): mewakili pertanyaan atau prediksi selanjutnya. Simpul anak dapat berupa:
 - Simpul keputusan (Decision node): mengajukan pertanyaan lanjutan berdasarkan fitur data.
 - Simpul daun (Leaf node): mewakili prediksi akhir.

Proses pembuatan pohon keputusan melibatkan pemisahan data menjadi subset berdasarkan fitur yang paling relevan. Pemisahan ini dilakukan secara berulang hingga mencapai kriteria pemberhentian yang telah ditentukan sebelumnya. Kriteria umum untuk pemberhentian meliputi:

- Semua data pada simpul memiliki kelas yang sama (homogen).
- Tidak ada lagi fitur yang relevan untuk pemisahan data.
- Mencapai kedalaman pohon yang maksimum.

Gini impurity adalah metrik yang digunakan untuk mengukur ketidakmurnian (atau ketidakpastian) suatu dataset. Dalam konteks pohon keputusan, Gini impurity digunakan untuk menentukan bagaimana data dibagi menjadi subset yang lebih homogen. Langkah-langkah membangun pohon keputusan:

1. Hitung Gini impurity untuk seluruh dataset
2. Hitung Gini impurity untuk setiap pemisahan berdasarkan variabel.
3. Pilih pemisahan dengan Gini impurity terendah.
4. Ulangi langkah 2 dan 3 untuk setiap subset data yang dihasilkan dari pemisahan pertama.
5. Tetapkan prediksi untuk setiap simpul daun (akhir cabang) berdasarkan kelas mayoritas data pada simpul tersebut.

Overfitting adalah masalah yang sering terjadi pada model pembelajaran mesin, termasuk pohon keputusan. Overfitting terjadi ketika model terlalu fokus pada detail data pelatihan dan tidak dapat digeneralisasi dengan baik ke data baru. Untuk mencegah overfitting pada pohon Keputusan diperlukan lebih banyak data pelatihan. Semakin banyak data yang digunakan, semakin kecil kemungkinan model untuk overfitting. Lalu Melakukan Pemangkasan Pohon. Pemangkasan pohon menghilangkan cabang-cabang yang tidak relevan pada pohon keputusan. Mencegah overfitting adalah kunci untuk membangun pohon keputusan yang akurat.