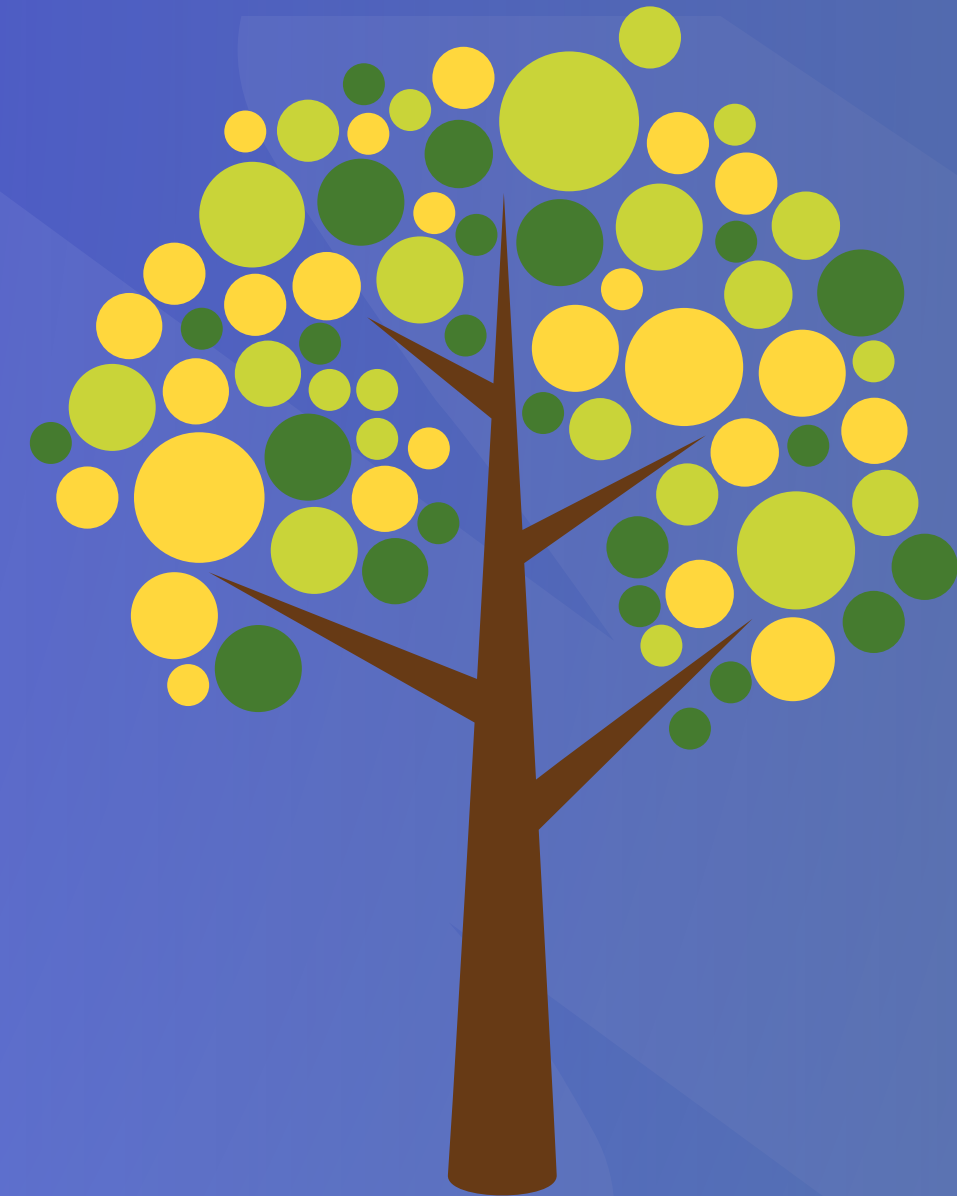


TI 5B

Decision Tree

▶ ▶ KELOMPOK 4

- 1. Amanda Rizky Yorina Prayoga
- 2. Dayinta Ayu Faj'rin
- 3. Diaz Azkha Varissa
- 4. Fadillah Dwi Anggraini



Pengertian Decission Tree

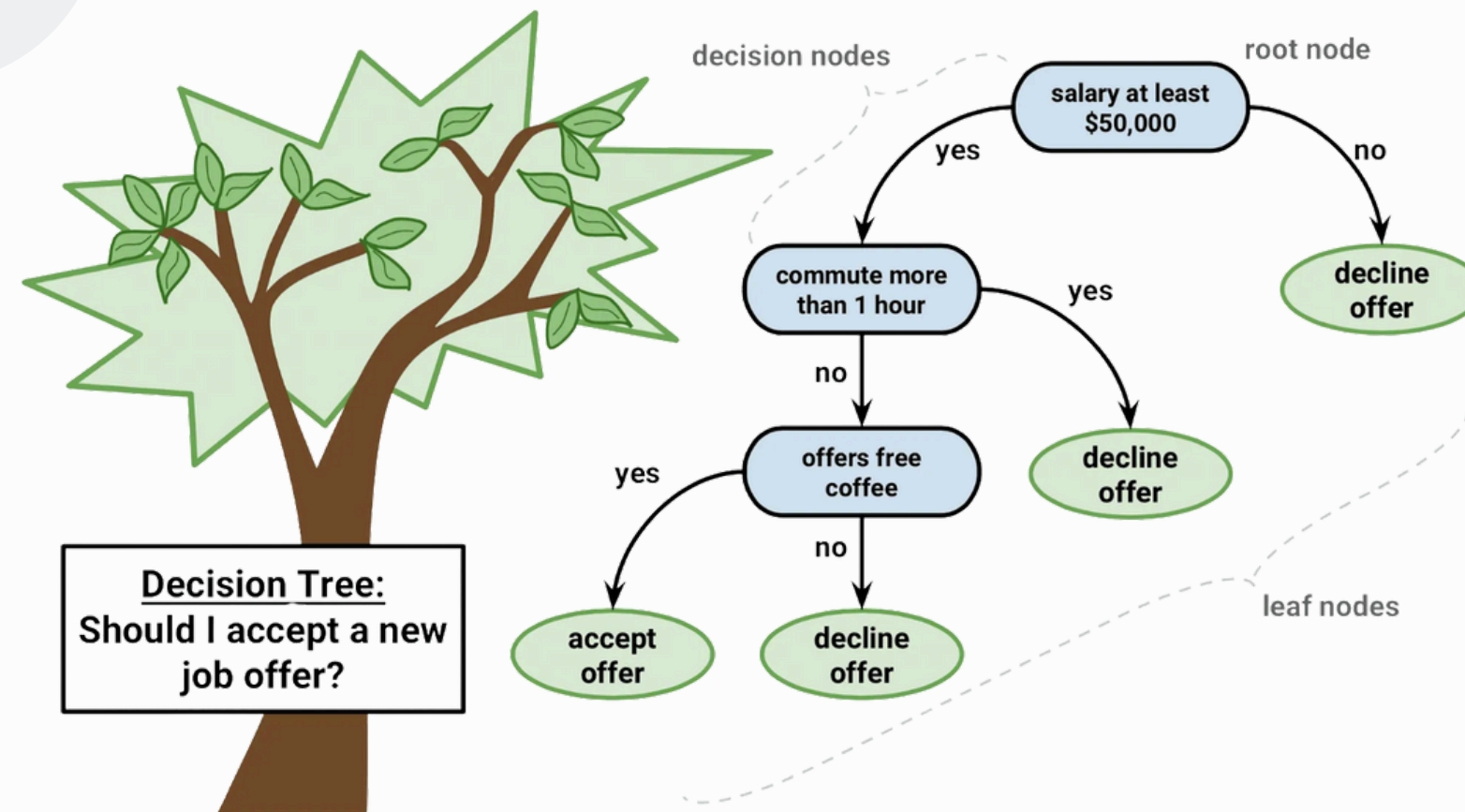
Decision tree adalah salah satu teknik klasifikasi yang populer dalam data mining. Metode decision tree ini menggunakan representasi pohon dimana setiap simpul atau node menggambarkan atribut, setiap cabang mewakili nilai analisa.

Dalam penerapannya, algoritma decision tree memerlukan atribut kelas karena masuk ke dalam kategori supervised learning. Oleh karena itu, data latih harus cukup besar dan beragam untuk memperoleh hasil yang optimal.

Decision Tree



Komponen Utama Decision Tree



Root Node

Titik awal dari pohon keputusan, yang memecah dataset berdasarkan fitur pertama

Decision Node

Titik percabangan dalam pohon keputusan di mana dataset dibagi lebih lanjut berdasarkan fitur tertentu.

Edges

Garis yang menghubungkan simpul-simpul dalam pohon, yang menunjukkan keputusan yang diambil.

Leaf Node

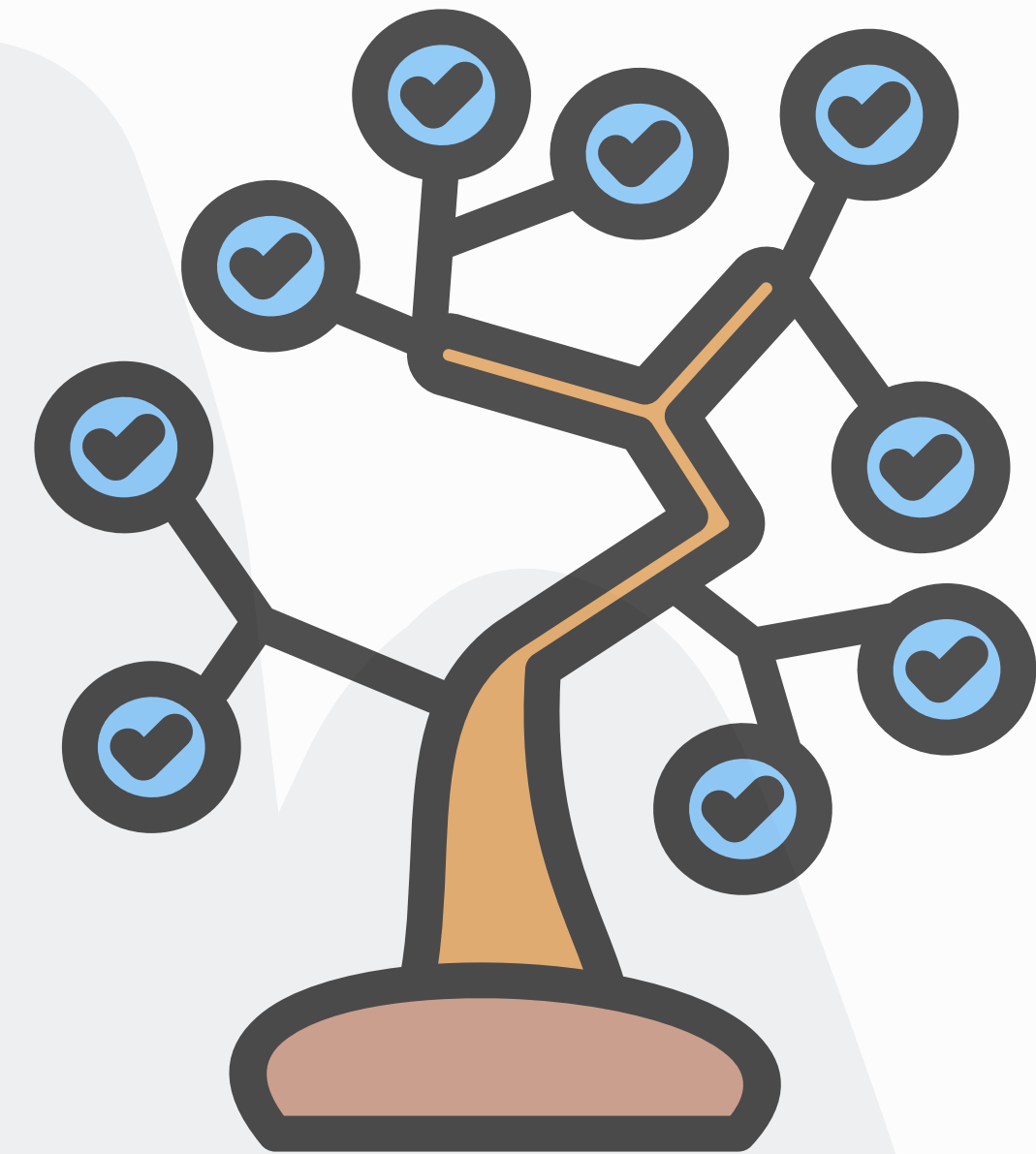
Simpul akhir yang menunjukkan label kelas atau nilai yang diprediksi.

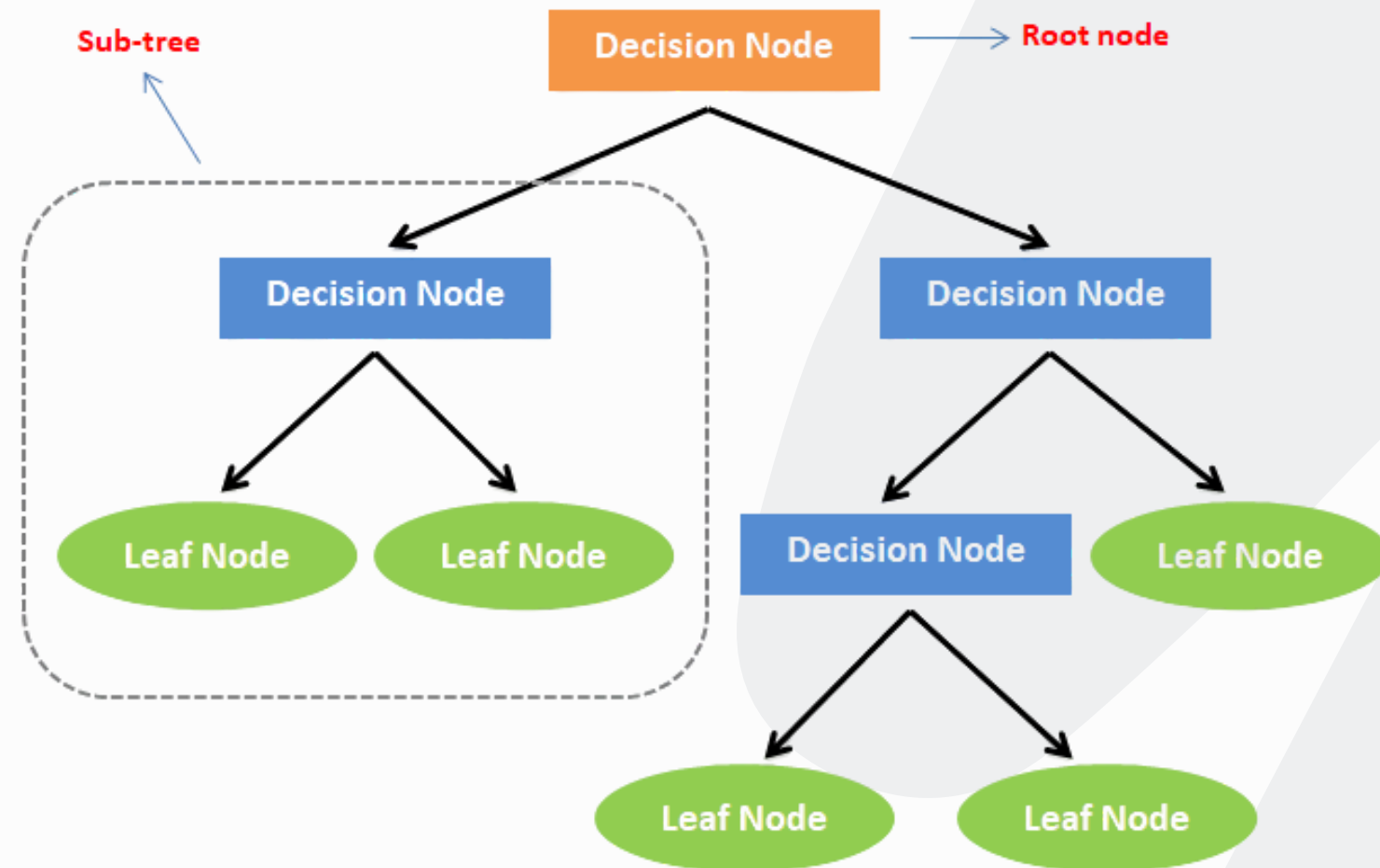


HIPOTESA FUNCTION

Hipotesa function pada Decision Tree adalah fungsi atau aturan yang digunakan model untuk memprediksi output (kelas atau nilai) berdasarkan input / pembelahan yang dilakukan. Hypothesis function mewakili “hipotesa” model mengenai bagaimana fitur-fitur dalam data menentukan hasil akhir.

Pada Decision Tree, hypothesis function bukan berupa persamaan matematis, melainkan serangkaian aturan (if-else) yang terbentuk sepanjang jalur dari root node ke leaf node.





Konsep hypothesis function adalah memetakan data input ke output dengan mengikuti struktur pohon keputusan. Setiap node di pohon memeriksa atribut tertentu, dan berdasarkan hasilnya, memilih cabang yang sesuai. Proses ini dilakukan hingga model mencapai leaf node, yang berisi hasil prediksi.

Secara umum, konsepnya yaitu:

1. Mulai dari root node.
2. Evaluasi kondisi pada node tersebut.
3. Ikuti cabang sesuai hasil kondisi.
4. Lanjutkan hingga mencapai leaf node.
5. Leaf node memberikan hasil prediksi → inilah output hypothesis function.

Dengan konsep ini, setiap prediksi pada Decision Tree dihasilkan dengan mengikuti urutan aturan yang terstruktur.

Cost Function

Terdapat beberapa cara untuk mengukur cost function, yaitu:

1. Gini Impurity Gini Gain

Mengukur ketidakmurnian atau keberagaman data dalam sebuah node. Semakin rendah nilai Gini, semakin baik pembelahan / pemisahan tersebut.

$$I_G(t) = 1 - \sum_{i=1}^c (p_i)^2$$

dengan keterangan rumus :

IG(t): Impurity Gini pada node t.

c: Jumlah kelas (misalnya, 2 kelas: "Ya" atau "Tidak").

pi: Proporsi (probabilitas) pengamatan yang termasuk dalam kelas i di node t.

Node yang murni (semua observasi kelasnya sama) akan memiliki IG = 0.

2. Entropy Information Gain

Mengukur ketidakpastian dalam suatu data atau node. Entropy semakin rendah menunjukkan bahwa data dalam node tersebut lebih homogen.

$$H(t) = - \sum_{i=1}^c p_i \log_2(p_i)$$

- H(t): Entropi pada node t.

- c: Jumlah kelas.

- pi: Proporsi pengamatan yang termasuk dalam kelas i di node t.

Node yang murni memiliki H = 0. Node yang paling tidak murni (kelasnya terdistribusi merata) memiliki H = 1 (untuk 2 kelas).



Pembelahan terbaik adalah yang meminimalkan nilai cost function (Gini atau Entropy), yang berarti data dalam masing-masing node akan lebih homogen (semua sample memiliki kelas yang sama).

21112015

Friday

Nov 2025

Thank you