

# Data Mining

## Decision Tree

# Definisi

- Sistem pendukung keputusan yang menggunakan model keputusan seperti pohon dan kemungkinan konsekuensinya, termasuk hasil acara kebetulan, biaya sumber daya, dan utilitas.

# Kontrol Bersyarat

- Ini adalah salah satu cara untuk menampilkan algoritma yang hanya berisi pernyataan kontrol bersyarat.

# Operational Research & Machine Learning

- Biasa digunakan dalam Operational Research, khususnya dalam analisis keputusan, untuk membantu mengidentifikasi strategi yang paling mungkin untuk mencapai tujuan, tetapi juga merupakan alat yang populer dalam Machine Learning.

# Struktur pohon untuk klasifikasi

- Memiliki struktur seperti *flow chart* di mana setiap simpul internal mewakili "tes" pada atribut, masing-masing cabang mewakili hasil pengujian, dan setiap simpul daun mewakili label kelas.
- Jalur dari root ke leaf mewakili aturan klasifikasi.

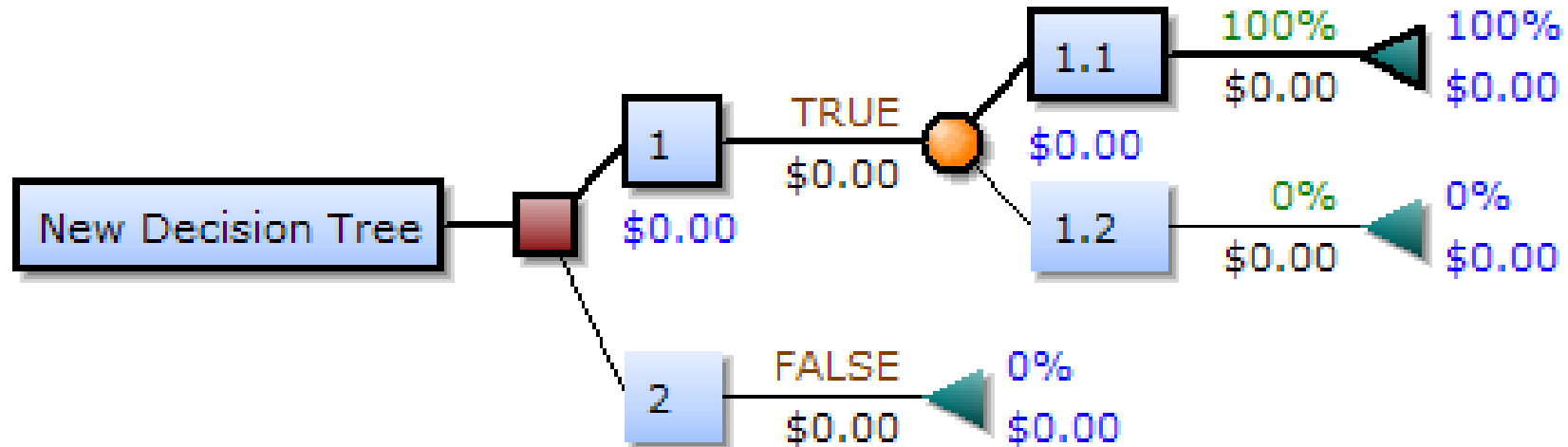
# Analisis keputusan visual

- Dalam analisis keputusan, pohon keputusan dan diagram pengaruh yang terkait erat digunakan sebagai alat pendukung keputusan visual dan analitis, di mana nilai yang diharapkan (atau utilitas yang diharapkan) dari alternatif yang bersaing dihitung.

# Jenis Nodes

- Pohon keputusan terdiri dari tiga jenis simpul:
  - Decision nodes - biasanya diwakili oleh kuadrat
  - Chance nodes - biasanya diwakili oleh lingkaran
  - End node - biasanya diwakili oleh segitiga

# Bentuk umum





# Splitting & Converging Path

- Diambil dari kiri ke kanan, pohon keputusan hanya memiliki node burst (*Splitting Path*) tetapi tidak ada node sink (*Converging Path*).
- Oleh karena itu, digunakan secara manual, mereka dapat tumbuh sangat besar dan seringkali sulit untuk menggambar sepenuhnya dengan tangan.

# Decision rules

- Pohon keputusan dapat dilinearisasi menjadi aturan keputusan, di mana hasilnya adalah isi simpul daun, dan kondisi di sepanjang jalur membentuk konjungsi dalam klausa if
  - jika kondisi1 dan kondisi2 dan kondisi3 maka hasil.
- Aturan keputusan dapat dihasilkan dengan membuat *Association Rules* dengan variabel target di sebelah kanan.

# Contoh lain

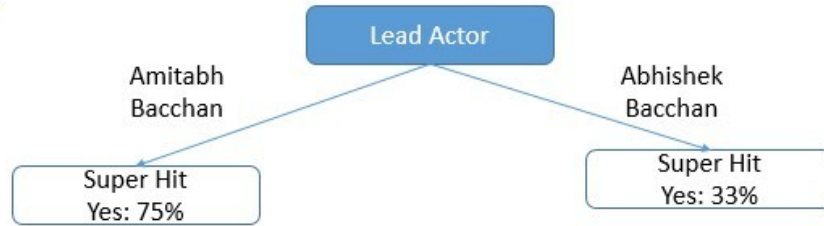


# Penjelasan

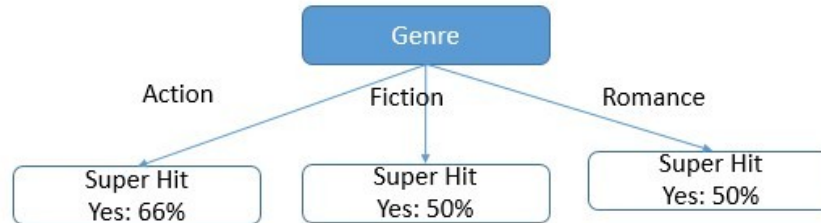
- Pohon yang ditunjukkan di atas membagi data sedemikian rupa sehingga diperoleh informasi maksimal, untuk memahami pohon :
  - Jika panjang rambut seseorang kurang dari 5 Inches, berat lebih dari 55 KG maka ada 80% peluang bagi orang tersebut menjadi Pria.

# Melihat dari berbagai sudut pandang

Method 1



Method 2



# Pembagian terbaik

- Dapat dengan jelas mengamati bahwa Metode 1 (Berdasarkan aktor utama) membagi data terbaik.
- Sedangkan metode kedua (Berdasarkan Genre) telah menghasilkan hasil yang beragam.

# Kelebihan Decision Tree (1)

- Sederhana untuk dipahami dan ditafsirkan. Manusia dapat memahami model pohon keputusan setelah penjelasan singkat.
- Memiliki nilai bahkan dengan sedikit data keras. Wawasan penting dapat dihasilkan berdasarkan para ahli yang menggambarkan situasi (alternatifnya, probabilitas, dan biaya) dan preferensi mereka untuk hasil.

# Kelebihan Decision Tree (2)

- Bantu menentukan nilai terburuk, terbaik, dan yang diharapkan untuk berbagai skenario.
- Gunakan model kotak putih. Jika hasil yang diberikan disediakan oleh model.
- Dapat dikombinasikan dengan teknik pengambilan keputusan lainnya.



# Kekurangan Decision Tree (1)

- Tidak stabil, yang berarti bahwa perubahan kecil dalam data dapat menyebabkan perubahan besar dalam struktur pohon keputusan yang optimal.
- Seringkali relatif tidak akurat. Banyak prediktor lain berkinerja lebih baik dengan data serupa. Ini dapat diperbaiki dengan mengganti pohon keputusan tunggal dengan hutan pohon keputusan acak, tetapi hutan acak tidak semudah ditafsirkan sebagai pohon keputusan tunggal.

# Kekurangan Decision Tree (2)

- Untuk data termasuk variabel kategorikal dengan jumlah level yang berbeda, perolehan informasi dalam pohon keputusan cenderung mendukung atribut tersebut dengan level yang lebih banyak.
- Perhitungan bisa menjadi sangat kompleks, terutama jika banyak nilai tidak pasti dan / atau jika banyak hasil dikaitkan.