



Teori dan Aplikasi Decision Tree (Pohon Keputusan)

Tim Pengajar Mata Kuliah Kecerdasan Buatan Tahun 2021





Outline

- Decision Tree
 - Contoh
 - Konsep
 - Entropi
 - Information gain
 - Demo
-





Decision Tree

- Alat pendukung keputusan yang menggunakan model keputusan seperti pohon dan kemungkinan konsekuensinya.
 - Termasuk supervised learning, dapat menyelesaikan regresi dan klasifikasi
 - An Inductive learning task
 - Menggunakan fakta untuk membuat kesimpulan yang lebih umum
 - Prediktif model berdasarkan serangkaian pengujian Boolean yang bercabang
 - Pengujian boolean lebih sederhana dibandingkan dengan satu tahapan pada classifier
-





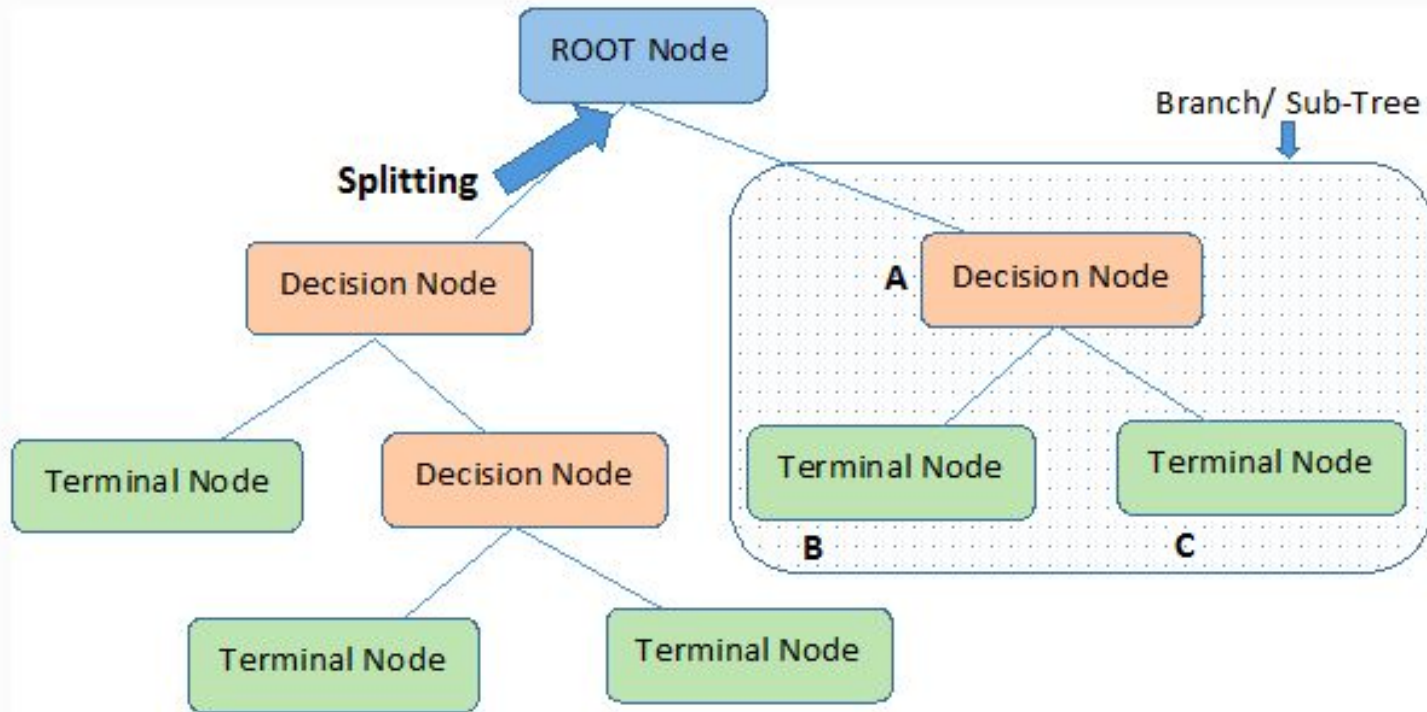
Istilah pada Decision Tree

- **Root Node:** merepresentasikan seluruh populasi
 - **Leaf Node/terminal Node:** node yang tidak dapat dipisahkan lagi menjadi node yang lain
 - **Parent/Child Node:** root node adalah parent node, dan semua cabangnya adalah child node
 - **Splitting:** Memisah root node/sub node menjadi bagian yang berbeda
 - **Branch/Subtree:** Dibentuk dengan splitting tree/node
 - **Pruning:** menghapus branch dari tree (kebalikan dari splitting)
-





Deskripsi Istilah pada Decision Tree

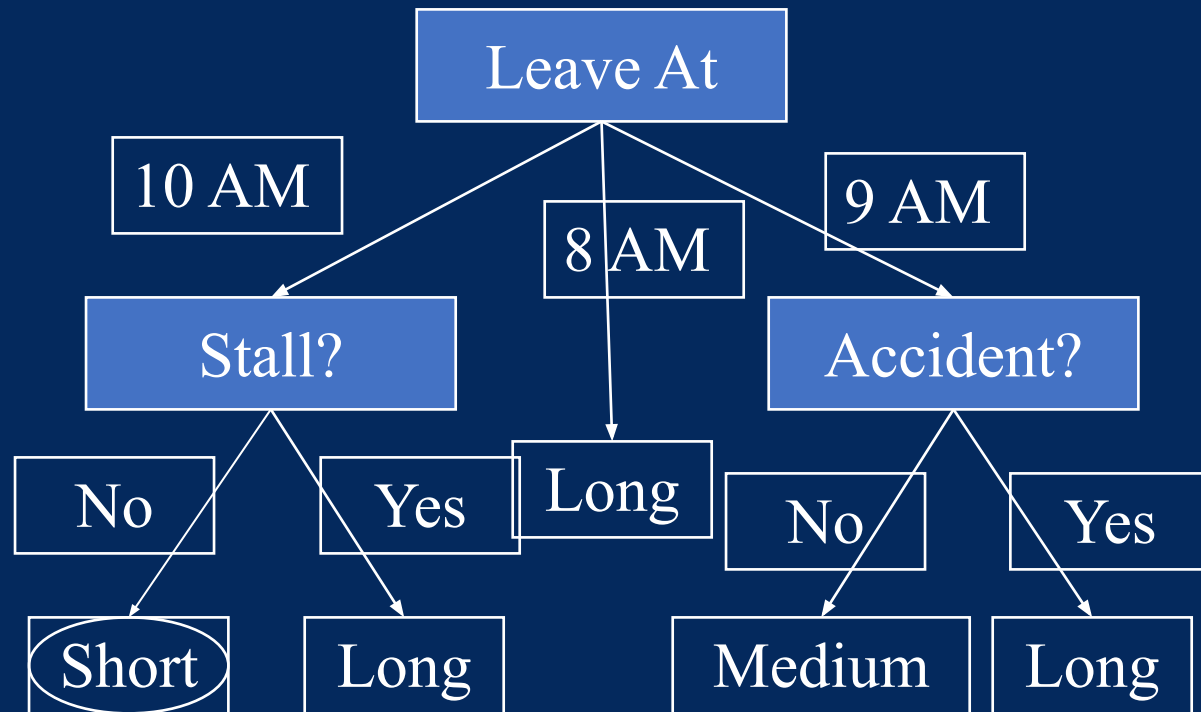


Note:- A is parent node of B and C.





Memprediksi Waktu Perjalanan



- Jika kita berangkat (leave) jam 10 pagi dan tidak ada mobil yang mogok (stall) di jalan, berapa waktu perjalanan kita?





Inductive Learning

- Pada decision tree, rangkaian keputusan dibuat secara Boolean dan mengikuti cabang yang sesuai
 - Apakah kita berangkat (leave) jam 10 pagi?
 - Apakah ada mobil yang mogok (Stall) di jalan?
 - Apakah ada kecelakaan (accident) di jalan raya?
 - Dengan menjawab setiap pertanyaan ya/tidak (Boolean) tersebut, kemudian sampai pada kesimpulan tentang berapa lama waktu perjalanan
-





Decision Tree sebagai Aturan

- Decision tree dapat direpresentasikan tidak dalam bentuk grafis.
 - Dapat direpresentasikan sebagai sekumpulan aturan. Namun, risikonya adalah jauh lebih sulit untuk dibaca.
-





Decision Trees sebagai Aturan

```
if hour == 8am
    commute time = long
else if hour == 9am
    if accident == yes
        commute time = long
    else
        commute time = medium
else if hour == 10am
    if stall == yes
        commute time = long
    else
        commute time = short
```

- Perhatikan bahwa semua atribut tidak harus digunakan di setiap jalur keputusan.
- Seperti pada kasus jika berangkat pukul 8 am. Atribut accident dan stall tidak muncul

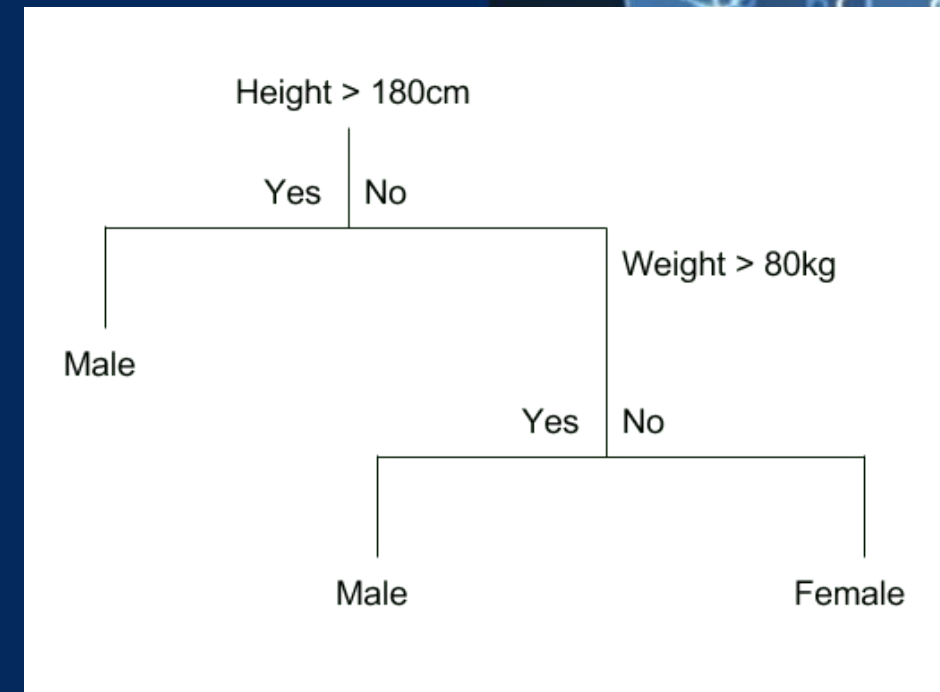




Decision Trees sebagai Aturan

```
1 If Height > 180 cm Then Male
2 If Height <= 180 cm AND Weight > 80 kg Then Male
3 If Height <= 180 cm AND Weight <= 80 kg Then Female
4 Make Predictions With CART Models
```

- The tree can be stored to file as a graph or a set of rules





Cara Membuat Decision Tree

- Membuat daftar atribut yang dapat diukur.
- Tentukan atribut yang akan diprediksi.
- Membuat tabel experience yang mendata apa yang sudah diamati.





Contoh Tabel

Example	Attributes				Target
	Hour	Weather	Accident	Stall	Commute
D1	8 AM	Sunny	No	No	Long
D2	8 AM	Cloudy	No	Yes	Long
D3	10 AM	Sunny	No	No	Short
D4	9 AM	Rainy	Yes	No	Long
D5	9 AM	Sunny	Yes	Yes	Long
D6	10 AM	Sunny	No	No	Short
D7	10 AM	Cloudy	No	No	Short
D8	9 AM	Rainy	No	No	Medium
D9	9 AM	Sunny	Yes	No	Long
D10	10 AM	Cloudy	Yes	Yes	Long
D11	10 AM	Rainy	No	No	Short
D12	8 AM	Cloudy	Yes	No	Long
D13	9 AM	Sunny	No	No	Medium





Memilih Atribut

- Pada tabel terdapat 4 atribut:
 - hour, weather, accident dan stall
 - Akan tetapi decision tree hanya menggunakan 3 atribut:
 - hour, accident dan stall
 - Bagaimana hal ini bisa terjadi?
-





Memilih Atribut

- Algoritma untuk memilih atribut yaitu ID3, C4.5, CART, CHAID, MARS.
 - Struktur dasar pembuatan decision tree sebagian besar sama.
 - Perbedaannya terletak pada cara memilih atribut.
 - Fokus pada algoritma ID3 yang dikembangkan oleh Ross Quinlan (1975).
-





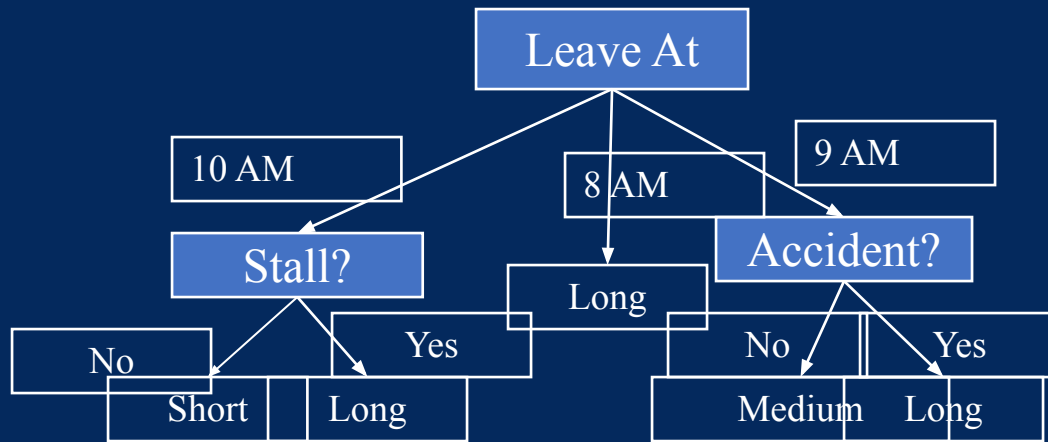
Algoritma Decision Tree

- Ide dasar Decision Tree adalah:
 - Pilih atribut terbaik untuk memisahkan data/sampel yang tersisa dan jadikan atribut itu simpul keputusan
 - Ulangi proses secara rekursif untuk setiap anak (child)
 - Hentikan ketika:
 - Semua data/sampel memiliki atribut target yang sama
 - Jika tidak ada lagi atribut
 - Jika tidak ada lagi data/sampel
-





Menentukan Atribut Terbaik



- Dari Decision Tree di atas, bagaimana dapat tahu untuk split/memisahkan pada *leave stall* dan *accident* tetapi tidak pada *weather*?





Langkah Algoritma ID3

1. Tree dimulai dengan sebuah simpul yang merepresentasikan sampel data pelatihan yaitu membuat node.
 2. Pada setiap iterasi pada atribut yang tidak digunakan hitung **entropi** dan **information gain**.
 3. Memilih atribut yang memiliki **entropi terkecil** atau **information gain terbesar**.
 4. Kemudian sample data di-split berdasarkan atribut terpilih.
 5. Algoritma terus berulang untuk setiap subset, memproses atribut yang belum dipilih pada langkah sebelumnya.
-





Algoritma ID3

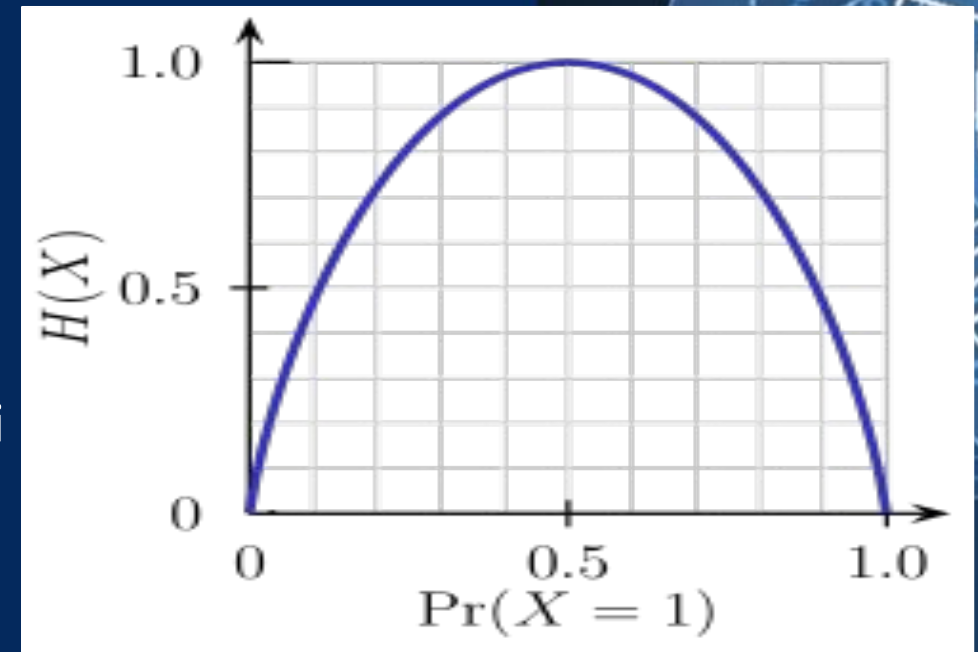
- Algoritma ID3 tidak pernah melakukan backtracking untuk merevisi keputusan pemilihan attribute yang telah dilakukan sebelumnya.
 - ID3 hanya menangani nilai-nilai attribute yang sedikit dan diskrit, tetapi algoritma modifikasinya, [algoritma C4.5](#) (1993) mampu menangani nilai attribute kontinu.
-





Entropi

- Entropi mengukur randomness (keacakan) dari informasi yang diproses, atau mengukur seberapa informatif sebuah node.
- Semakin tinggi nilai entropi maka semakin susah untuk membuat kesimpulan dari informasi tersebut.
- Melempar koin adalah contoh yang menghasilkan informasi yang random.
- Dari grafik dapat diamati entropi $H(X)$ bernilai 0 ketika probability bernilai 0 atau 1. Entropi bernilai maksimum ketika probabilitasnya 0.5 karena keacakannya optimal.





Entropi

- Pada ID3 – branch dengan entropi bernilai 0 menjadi **leaf node**.
- Branch dengan entropi lebih dari 0 (>0) perlu untuk **di-split**





Entropi

- Contoh lain:
- Entropi diminimalkan jika semua nilai atribut target sama.
 - Jika tahu bahwa waktu **perjalanan akan selalu singkat**, maka **entropi = 0**
- Entropi dimaksimalkan ketika nilai atribut target memiliki peluang yang sama (misalnya hasilnya acak)
 - Jika **commute time = short** untuk 3 sampel, **medium** untuk 3 sampel dan **long** untuk 3 sampel, entropy dimaksimalkan





Entropi

- Perhitungan entropi untuk 1 atribut
 - S = himpunan dataset kasus
 - c = banyaknya partsisi S
 - p_i = probabilitas dari event “ i ” dari S

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy (5,9)
= Entropy (0.36, 0.64)
= - (0.36 \log_2 0.36) - (0.64 \log_2 0.64)
= 0.94



Entropi

- Perhitungan entropi untuk banyak atribut.
 - T = State saat ini
 - X = Atribut yang dipilih

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

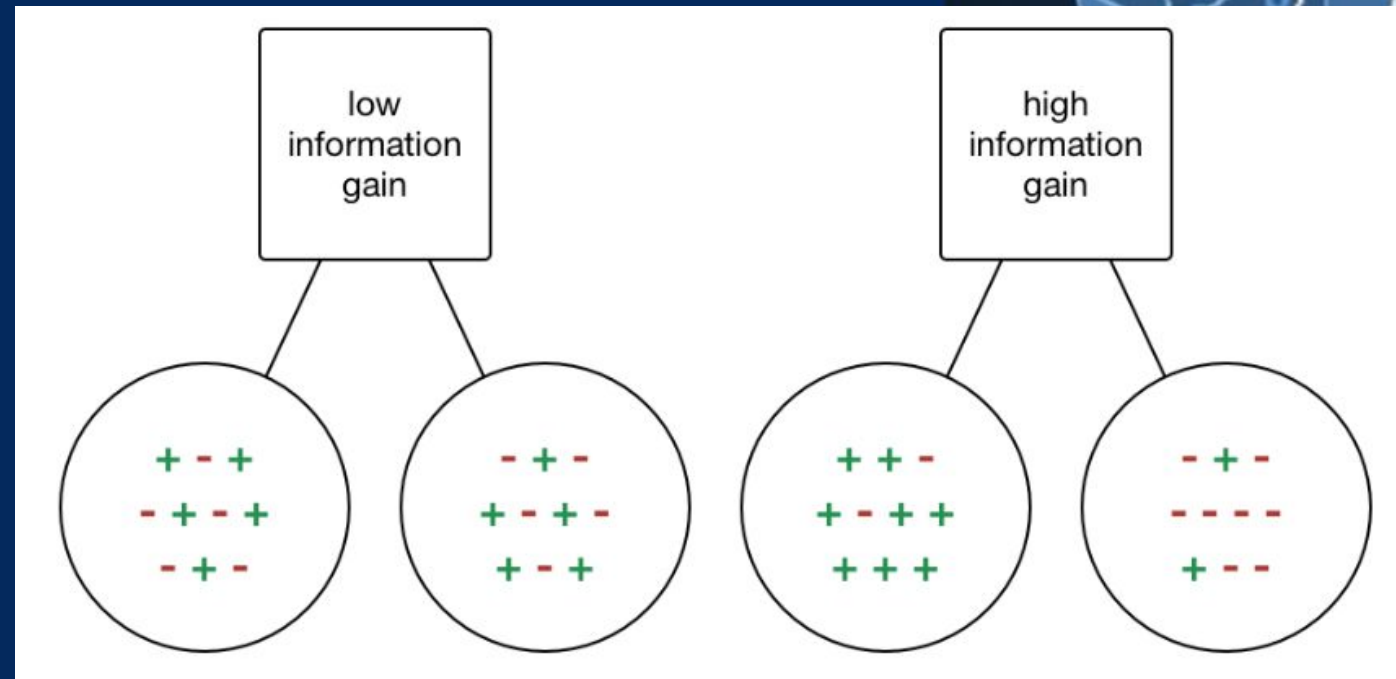


$$\begin{aligned} E(\text{PlayGolf, Outlook}) &= P(\text{Sunny}) \cdot E(3,2) + P(\text{Overcast}) \cdot E(4,0) + P(\text{Rainy}) \cdot E(2,3) \\ &= (5/14) \cdot 0.971 + (4/14) \cdot 0.0 + (5/14) \cdot 0.971 \\ &= 0.693 \end{aligned}$$



Information Gain

- Digunakan untuk mengukur seberapa tepat atribut memisahkan data training sesuai dengan targetnya.
- Membangun decision tree adalah mencari atribut yang memiliki **information gain yang tinggi** dan **entropi yang minimal**.





Information Gain

- Mengurangi nilai entropi
- Menghitung perbedaan entropi sebelum di split dan rata-rata entropi setelah di split
- Atau dapat ditulis seperti berikut:

$$\text{Information Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$

$$\begin{aligned}\text{IG}(\text{PlayGolf}, \text{Outlook}) &= E(\text{PlayGolf}) - E(\text{PlayGolf}, \text{Outlook}) \\ &= 0.940 - 0.693 \\ &= 0.247\end{aligned}$$

$$\text{Information Gain} = \text{Entropy}(\text{before}) - \sum_{j=1}^K \text{Entropy}(j, \text{after})$$



Masalah pada Decision Tree

- Decision tree dapat mengklasifikasi dengan cepat, akan tetapi waktu untuk membuat tree paling tinggi dibanding dengan classifier yang lain.
- Decision tree jika terdapat masalah dapat mempengaruhi seluruh tree
 - Dapat menjadi masalah yang serius ketika jumlah kelas meningkat





Contoh 1

- Data yang ada pada Tabel akan digunakan untuk membentuk decision tree.
- Terdiri dari atribut-atribut Cuaca, Suhu, Kelembaban, dan Berangin.
- Kelasnya ada pada kolom Target yaitu kelas “**Tidak**” dan kelas “**Ya**”.

No	Cuaca	Suhu	Kelembaban	Berangin	Main
1	Cerah	Panas	Tinggi	Tidak Kencang	Tidak
2	Cerah	Panas	Tinggi	Kencang	Tidak
3	Berawan	Panas	Tinggi	Tidak Kencang	Ya
4	Hujan	Sejuk	Tinggi	Tidak Kencang	Ya
5	Hujan	Dingin	Normal	Tidak Kencang	Ya
6	Hujan	Dingin	Normal	Kencang	Tidak
7	Berawan	Dingin	Normal	Kencang	Ya
8	Cerah	Sejuk	Tinggi	Tidak Kencang	Tidak
9	Cerah	Dingin	Normal	Tidak Kencang	Ya
10	Hujan	Sejuk	Normal	Tidak Kencang	Ya
11	Cerah	Sejuk	Normal	Kencang	Ya
12	Berawan	Sejuk	Tinggi	Kencang	Ya
13	Berawan	Panas	Normal	Tidak Kencang	Ya
14	Hujan	Sejuk	Tinggi	Kencang	Tidak



Contoh 1

- Dari data tersebut dianalisis; dataset memiliki 14 kasus yang terdiri 9 “Ya” dan 5 “Tidak” pada kolom Main.
- $Entropi (S) = -(9/14) \times \log_2 (9/14) + -(5/14) \times \log_2 (5/14)$
= 0,9403

Total Kasus	Ya	Tidak	Entropi
14	9	5	0,9403





Contoh 1 – Langkah 2

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

- Menghitung entropi (dengan formula multiple atribut) dan information gain setiap atribut
- $E(\text{main, cuaca}) = P(\text{cerah}) * E(2,3) + P(\text{hujan}) * E(3,2) + P(\text{berawan}) * E(4,0) = P(5/14) * E(2,3) + P(5/14) * E(3,2) + P(4/14) * E(4,0)$
- Untuk menyelesaikan persamaan di atas, cari terlebih dahulu entropi setiap kondisi
- $E_{\text{cerah}}(p,n) = E(\text{jumlah cerah dan Ya, jumlah cerah dan tidak})$
- $E_{\text{cerah}}(2,3) = -(2/(2+3))\log_2(2/(2+3)) - (3/(2+3))\log_2(3/(2+3)) = 0,9710$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
1	Cuaca	Berawan	4	4	0	0,0000	
		Hujan	5	3	2	0,9710	
		Cerah	5	2	3	0,9710	
		Weight Sum Entropy	14			0,6935	0,2467
	Suhu	Dingin	4	3	1	0,8113	
		Panas	4	2	2	1,0000	
		Sejuk	6	4	2	0,9183	
		Weight Sum Entropy	14			0,9111	0,0292
	Kelembaban	Tinggi	7	3	4	0,9852	
		Normal	7	6	1	0,591672779	
		Weight Sum Entropy	14			0,7885	0,1518
	Berangin	Tidak Kencang	8	6	2	0,8113	
		Kencang	6	3	3	1,0000	
		Weight Sum Entropy	14			0,8922	0,0481



Contoh 1 – Langkah 2

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

- $E_{\text{berawan}}(p,n) = E(\text{jumlah berawan dan Ya, jumlah berawan dan tidak})$
- $E_{\text{berawan}}(4,0) = -(4/(4+0)) \log_2(4/(4+0)) + -(0/(4+0)) \log_2(0/(4+0)) = 0$
- $E_{\text{hujan}}(p,n) = E(\text{jumlah hujan dan Ya, jumlah hujan dan tidak})$
- $E_{\text{hujan}}(3,2) = -(3/(3+2)) \log_2(3/(3+2)) + -(2/(3+2)) \log_2(2/(3+2)) = 0,970950594$
- Selesaikan persamaan sebelumnya
- $E(\text{main, cuaca}) = P(5/14) * 0.9710 + P(5/14) * 0.9710 + P(4/14) * 0 = 0,6935$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
1	Cuaca	Berawan	4	4	0	0,0000	0,2467
		Hujan	5	3	2	0,9710	
		Cerah	5	2	3	0,9710	
		Weight Sum Entropy	14			0,6935	
	Suhu	Dingin	4	3	1	0,8113	0,0292
		Panas	4	2	2	1,0000	
		Sejuk	6	4	2	0,9183	
		Weight Sum Entropy	14			0,9111	
	Kelembaban	Tinggi	7	3	4	0,9852	0,1518
		Normal	7	6	1	0,591672779	
		Weight Sum Entropy	14			0,7885	
	Berangin	Tidak Kencang	8	6	2	0,8113	0,0481
		Kencang	6	3	3	1,0000	
		Weight Sum Entropy	14			0,8922	



Contoh 1 – Langkah 2

$$\text{Information Gain} = \text{Entropy}(\text{before}) - \sum_{j=1}^K \text{Entropy}(j, \text{after})$$

- Information gain dari atribut cuaca = entropi sebelum – entropi setelah split
- Information gain cuaca = 0,9403 – 0,6935 = **0,2467**
- Hitung entropi dan information gain untuk setiap atribut.

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
1	Cuaca	Berawan	4	4	0	0,0000	
		Hujan	5	3	2	0,9710	
		Cerah	5	2	3	0,9710	
		Weight Sum Entropy	14			0,6935	0,2467
	Suhu	Dingin	4	3	1	0,8113	
		Panas	4	2	2	1,0000	
		Sejuk	6	4	2	0,9183	
		Weight Sum Entropy	14			0,9111	0,0292
	Kelembaban	Tinggi	7	3	4	0,9852	
		Normal	7	6	1	0,591672779	
		Weight Sum Entropy	14			0,7885	0,1518
	Berangin	Tidak Kencang	8	6	2	0,8113	
		Kencang	6	3	3	1,0000	
		Weight Sum Entropy	14			0,8922	0,0481



Contoh 1 – Langkah 2

$$Information\ Gain = Entropy(before) - \sum_{j=1}^K Entropy(j, after)$$

$$E(T, X) = \sum_{c \in X} P(c) E(c)$$

- Hitung entropi dan information gain untuk setiap atribut.

E(Play Golf, Temperature)

$$E(panas) = - \left(\frac{2}{4} \log_2 \left(\frac{2}{4} \right) + \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right) = 1.0$$

$$E(sejuk) = - \left(\frac{4}{6} \log_2 \left(\frac{4}{6} \right) + \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right) = 0.9183$$

$$E(dingin) = - \left(\frac{3}{4} \log_2 \left(\frac{3}{4} \right) + \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right) = 0.8113$$

Gain = E(before) – ∑_{j=1} E(j, after)

$$= 0.94 - \left(\frac{4}{14} * 1 + \frac{6}{14} * 0.9183 + \frac{4}{14} * 0.8113 \right)$$

$$= 0.0292$$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
1	Cuaca	Berawan	4	4	0	0,0000	
		Hujan	5	3	2	0,9710	
		Cerah	5	2	3	0,9710	
		Weight Sum Entropy	14			0,6935	0,2467
	Suhu	Dingin	4	3	1	0,8113	
		Panas	4	2	2	1,0000	
		Sejuk	6	4	2	0,9183	
		Weight Sum Entropy	14			0,9111	0,0292
	Kelembaban	Tinggi	7	3	4	0,9852	
		Normal	7	6	1	0,591672779	
		Weight Sum Entropy	14			0,7885	0,1518
	Berangin	Tidak Kencang	8	6	2	0,8113	
		Kencang	6	3	3	1,0000	
		Weight Sum Entropy	14			0,8922	0,0481



Contoh 1 – Langkah 2

$$\text{Information Gain} = \text{Entropy}(\text{before}) - \sum_{j=1}^K \text{Entropy}(j, \text{after})$$
$$E(T, X) = \sum_{c \in X} P(c) E(c)$$

- Hitung entropi dan information gain untuk setiap atribut.

E(Play Golf, Humidity)

$$E(\text{tinggi}) = - \left(\frac{3}{7} \log_2 \left(\frac{3}{7} \right) + \frac{4}{7} \log_2 \left(\frac{4}{7} \right) \right) = 0.9852$$

$$E(\text{Normal}) = - \left(\frac{6}{7} \log_2 \left(\frac{6}{7} \right) + \frac{1}{7} \log_2 \left(\frac{1}{7} \right) \right) = 0.5916$$

Gain =

$$= 0.94 - \left(\frac{7}{14} * 1 + \frac{6}{14} * 0.9852 + \frac{1}{14} * 0.5916 \right) \\ = 0.1516$$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
1	Cuaca	Berawan	4	4	0	0,0000	
		Hujan	5	3	2	0,9710	
		Cerah	5	2	3	0,9710	
		Weight Sum Entropy	14			0,6935	0,2467
	Suhu	Dingin	4	3	1	0,8113	
		Panas	4	2	2	1,0000	
		Sejuk	6	4	2	0,9183	
		Weight Sum Entropy	14			0,9111	0,0292
	Kelembaban	Tinggi	7	3	4	0,9852	
		Normal	7	6	1	0,591672779	
		Weight Sum Entropy	14			0,7885	0,1518
	Berangin	Tidak Kencang	8	6	2	0,8113	
		Kencang	6	3	3	1,0000	
		Weight Sum Entropy	14			0,8922	0,0481



Contoh 1 – Langkah 2

$$Information\ Gain = Entropy(before) - \sum_{j=1}^K Entropy(j, after)$$

$$E(T, X) = \sum_{c \in X} P(c) E(c)$$

- Hitung entropi dan information gain untuk setiap atribut.

E(Play Golf, Windy)

$$E(tidak) = - \left(\frac{6}{8} \log_2 \left(\frac{6}{8} \right) + \frac{2}{8} \log_2 \left(\frac{2}{8} \right) \right) = 0.811$$

$$E(kencang) = - \left(\frac{3}{6} \log_2 \left(\frac{3}{6} \right) + \frac{3}{6} \log_2 \left(\frac{3}{6} \right) \right) = 1$$

Gain =

$$= 0.94 - \left(\frac{8}{14} * 0.811 + \frac{6}{14} * 1 \right)$$

$$= 0.0478$$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
1	Cuaca	Berawan	4	4	0	0,0000	
		Hujan	5	3	2	0,9710	
		Cerah	5	2	3	0,9710	
		Weight Sum Entropy	14			0,6935	0,2467
	Suhu	Dingin	4	3	1	0,8113	
		Panas	4	2	2	1,0000	
		Sejuk	6	4	2	0,9183	
		Weight Sum Entropy	14			0,9111	0,0292
	Kelembaban	Tinggi	7	3	4	0,9852	
		Normal	7	6	1	0,591672779	
		Weight Sum Entropy	14			0,7885	0,1518
	Berangin	Tidak Kencang	8	6	2	0,8113	
		Kencang	6	3	3	1,0000	
		Weight Sum Entropy	14			0,8922	0,0481



Contoh 1 – Langkah 3 dan 4

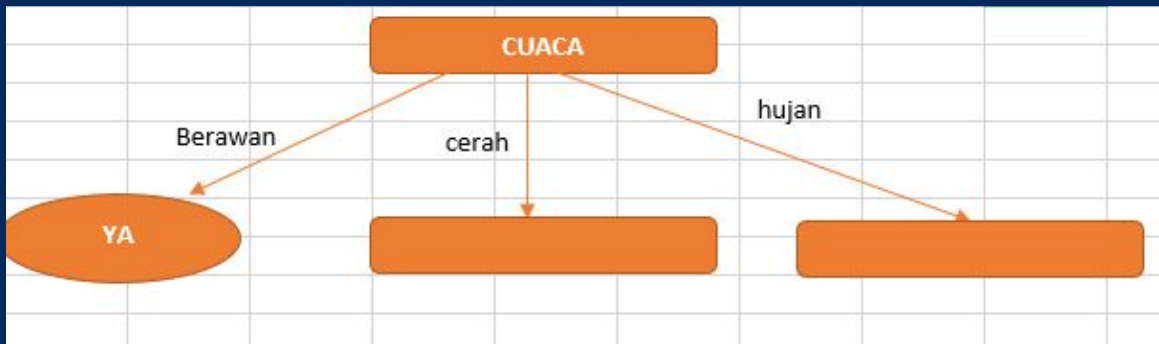
- Mencari atribut yang memiliki information gain paling tinggi, dan diperoleh atribut cuaca dengan nilai 0,246.
- **Node di-split berdasarkan atribut cuaca.**
- **Atribut berawan tidak perlu di-split karena nilai entropinya adalah 0.**
- Ulangi langkah 1 s.d. 4 untuk atribut yang lain.





Contoh 1

- **Atribut berawan tidak perlu di-split karena nilai entropinya adalah 0.**



No	Cuaca	Suhu	Kelembaban	Berangin	Main
3	Berawan	Panas	Tinggi	Tidak Kencan	Ya
7	Berawan	Dingin	Normal	Kencang	Ya
12	Berawan	Sejuk	Tinggi	Kencang	Ya
13	Berawan	Panas	Normal	Tidak Kencan	Ya



Contoh 1 – Atribut CeraH

- Komputasi untuk cabang cuaca dengan atribut CeraH

Total Kasus CeraH	Ya	Tidak	Entropi
5	2	3	0,97095059

No	Cuaca	Suhu	Kelembaban	Angin	Main
1	Cerah	Panas	Tinggi	Tidak Kencang	Tidak
2	Cerah	Panas	Tinggi	Kencang	Tidak
3	Cerah	Sejuk	Tinggi	Tidak Kencang	Tidak
4	Cerah	Dingin	Normal	Tidak Kencang	Ya
5	Cerah	Sejuk	Normal	Kencang	Ya



Contoh 1

• $E(\text{Play Golf, Temperature})$

- $E(\text{dingin}) = -\left(\frac{0}{1} \log_2 \left(\frac{0}{1}\right) + \frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) = 0$
- $E(\text{sejuk}) = -\left(\frac{1}{2} \log_2 \left(\frac{1}{2}\right) + \frac{1}{2} \log_2 \left(\frac{1}{2}\right)\right) = 1$
- $E(\text{dingin}) = -\left(\frac{0}{2} \log_2 \left(\frac{0}{2}\right) + \frac{2}{2} \log_2 \left(\frac{2}{2}\right)\right) = 0$
- **Gain** = $E(\text{before}) - \sum_{j=1} E(j, \text{after})$
- $= 0.9710 - \left(\frac{1}{5} * 0 + \frac{2}{5} * 0 + \frac{2}{5} * 1\right)$
- $= 0.5710$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
2	Suhu	Dingin	1	1	0	0,0000	
		Panas	2	0	2	0,0000	
		Sejuk	2	1	1	1,0000	
		Weight Sum Entropy	5			0,4000	0,5710
	Kelembaban	Tinggi	3	0	3	0	
		Normal	2	2	0	0	
		Weight Sum Entropy	5			0	0,9710
	Berangin	Tidak Kencang	3	1	2	0,9183	
		Kencang	2	1	1	1,0000	
		Weight Sum Entropy	5			0,9510	0,0200



Contoh 1

• $E(\text{Play Golf, Humidity})$

$$\bullet E(\text{tinggi}) = -\left(\frac{0}{3} \log_2 \left(\frac{0}{3}\right) + \frac{3}{3} \log_2 \left(\frac{3}{3}\right)\right) = 0$$

$$\bullet E(\text{Normal}) = -\left(\frac{2}{2} \log_2 \left(\frac{2}{2}\right) + \frac{0}{2} \log_2 \left(\frac{0}{2}\right)\right) = 0$$

• $\text{Gain} =$

$$\bullet = 0.9710 - \left(\frac{3}{5} * 0 + \frac{2}{5} * 0\right)$$

$$\bullet = 0.9710$$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
2	Suhu	Dingin	1	1	0	0,0000	
		Panas	2	0	2	0,0000	
		Sejuk	2	1	1	1,0000	
		Weight Sum Entropy	5			0,4000	0,5710
	Kelembaban	Tinggi	3	0	3	0	
		Normal	2	2	0	0	
		Weight Sum Entropy	5			0	0,9710
	Berangin	Tidak Kencang	3	1	2	0,9183	
		Kencang	2	1	1	1,0000	
		Weight Sum Entropy	5			0,9510	0,0200



Contoh 1

- $E(\text{Play Golf, Angin})$

- $E(\text{tidak}) = -\left(\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right) = 0,9183$

- $E(\text{kecang}) = -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) = 1$

- $\text{Gain} =$

- $= 0,9710 - \left(\frac{3}{5} * 0,9183 + \frac{2}{5} * 1\right)$

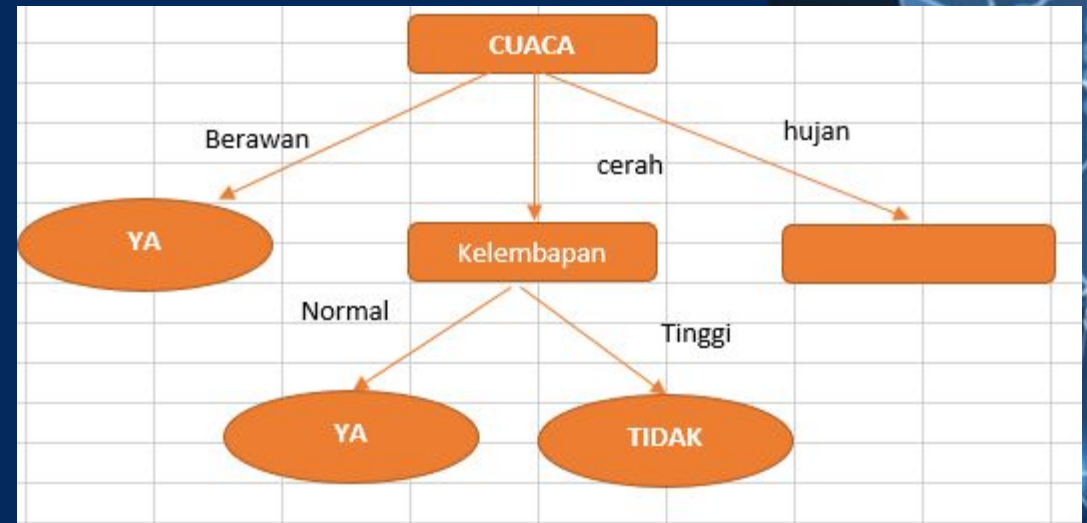
- $= 0,0200$

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
2	Suhu	Dingin	1	1	0	0,0000	
		Panas	2	0	2	0,0000	
		Sejuk	2	1	1	1,0000	
		Weight Sum Entropy	5			0,4000	0,5710
	Kelembaban	Tinggi	3	0	3	0	
		Normal	2	2	0	0	
		Weight Sum Entropy	5			0	0,9710
	Berangin	Tidak Kencang	3	1	2	0,9183	
		Kencang	2	1	1	1,0000	
		Weight Sum Entropy	5			0,9510	0,0200



Contoh 1

- Berdasarkan perhitungan maka cabang dari Cerah adalah **Kelembaban**





Contoh 2 – Atribut Hujan

- Komputasi untuk cabang cuaca dengan atribut Hujan

Total Kasus Hujan	Ya	Tidak	Entropi
5	3	2	0,97095059

- Lakukan perhitungan IG untuk masing-masing suhu, kelembapan, berangin

No	Cuaca	Suhu	Kelembaban	Berangin	Main
4	Hujan	Sejuk	Tinggi	Tidak Kencang	Ya
5	Hujan	Dingin	Normal	Tidak Kencang	Ya
6	Hujan	Dingin	Normal	Kencang	Tidak
10	Hujan	Sejuk	Normal	Tidak Kencang	Ya
14	Hujan	Sejuk	Tinggi	Kencang	Tidak

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Play Golf	
Yes	No
9	5

$$\begin{aligned} \text{Entropy(PlayGolf)} &= \text{Entropy}(5,9) \\ &= \text{Entropy}(0.36, 0.64) \\ &= -(0.36 \log_2 0.36) - (0.64 \log_2 0.64) \\ &= 0.94 \end{aligned}$$



Contoh 2

- Berdasarkan perhitungan maka cabang dari **hujan** adalah **Berangin**

Node	Atribut	Nilai	Sum(Nilai)	Sum(Ya)	Sum(Tidak)	Entropi	Gain
2	Suhu	Dingin	2	1	1	1,0000	
		Sejuk	3	2	1	0,918295834	
		Weight Sum Entropy	5				0,9509775
	Kelembaban	Tinggi	2	1	1	1,0000	
		Normal	3	2	1	0,918295834	
		Weight Sum Entropy	5				0,9509775
	Berangin	Tidak Kencang	3	3	0	0	
		Kencang	2	0	2	0	
		Weight Sum Entropy	5			0	0,9710



Hasil



No	Cuaca	Suhu	Kelembaban	Berangin	Main
1	Cerah	Panas	Tinggi	Tidak Kencang	Tidak
2	Cerah	Panas	Tinggi	Kencang	Tidak
3	Berawan	Panas	Tinggi	Tidak Kencang	Ya
4	Hujan	Sejuk	Tinggi	Tidak Kencang	Ya
5	Hujan	Dingin	Normal	Tidak Kencang	Ya
6	Hujan	Dingin	Normal	Kencang	Tidak
7	Berawan	Dingin	Normal	Kencang	Ya
8	Cerah	Sejuk	Tinggi	Tidak Kencang	Tidak
9	Cerah	Dingin	Normal	Tidak Kencang	Ya
10	Hujan	Sejuk	Normal	Tidak Kencang	Ya
11	Cerah	Sejuk	Normal	Kencang	Ya
12	Berawan	Sejuk	Tinggi	Kencang	Ya
13	Berawan	Panas	Normal	Tidak Kencang	Ya
14	Hujan	Sejuk	Tinggi	Kencang	Tidak



Additional Sources

- <https://medium.com/machine-learning-guy/an-introduction-to-decision-tree-learning-id3-algorithm-54c74eb2ad55>
- <https://towardsdatascience.com/decision-trees-for-classification-id3-algorithm-explained-89df76e72df1>





Kecerdasan Buatan JTI POLINEMA

- Terima Kasih -

