



ALGORITMA DECISION TREE C4.5

Nugroho DS

Pendahuluan

Sebagian besar keputusan yang dibuat manusia. Adanya informasi yang tidak sempurna atau imperfect information. Bagaimana memperoleh keputusan yang baik? Yaitu memilih berbagai alternatif yang berdasar perhitungan dengan nilai keputusan berbeda.

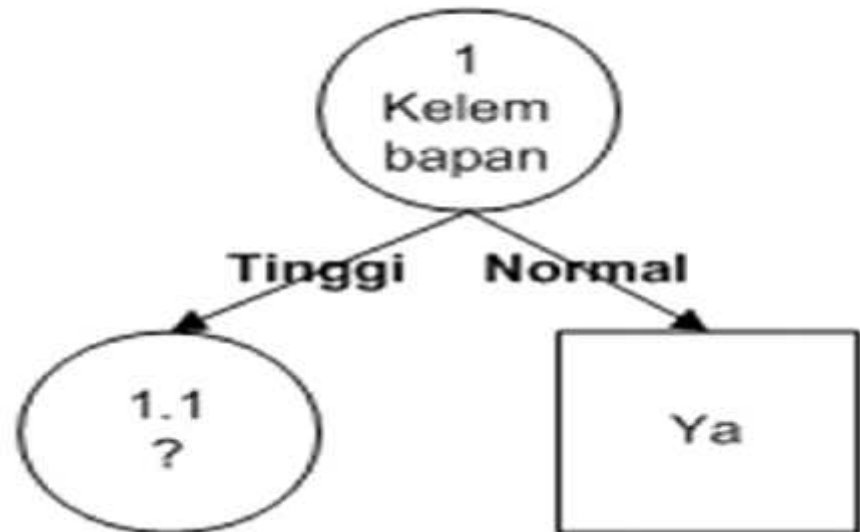
Decision Tree (Pohon Keputusan) merupakan salah satu alat untuk mencapai keputusan terbaik tersebut

Sejarah

- Salah satu algoritma induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3).
- ID3 dikembangkan oleh J. Ross Quinlan.
- Dalam prosedur algoritma ID3, input berupa sampel training, label training dan atribut.
- Algoritma C4.5 merupakan pengembangan dari ID3.

Algoritma C4.5

- algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif.
- Dasar algoritma C4.5 adalah pembentukan pohon keputusan (decision tree)



Rumu Algoritma C4.5

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) .$$

Dengan :

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots$$

dengan :

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

NO	CUACA	TEMPERATUR	KELEMBAPAN	BERANGIN	MAIN
1	Cerah	Panas	Tinggi	TIDAK	Tidak
2	Cerah	Panas	Tinggi	YA	Tidak
3	Berawan	Panas	Tinggi	TIDAK	Ya
4	Hujan	Sedang	Tinggi	TIDAK	Ya
5	Hujan	Dingin	Normal	TIDAK	Ya
6	Hujan	Dingin	Normal	YA	Ya
7	Berawan	Dingin	Normal	YA	Ya
8	Cerah	Sedang	Tinggi	TIDAK	Tidak
9	Cerah	Dingin	Normal	TIDAK	Ya
10	Hujan	Sedang	Normal	TIDAK	Ya
11	Cerah	Sedang	Normal	YA	Ya
12	Berawan	Sedang	Tinggi	YA	Ya
13	Berawan	Panas	Normal	TIDAK	Ya
14	Hujan	Sedang	Tinggi	YA	Tidak

$$\left(-\frac{4}{14} * \log_2\left(\frac{4}{14}\right)\right) + \left(-\frac{10}{14} * \log_2\left(\frac{10}{14}\right)\right)$$

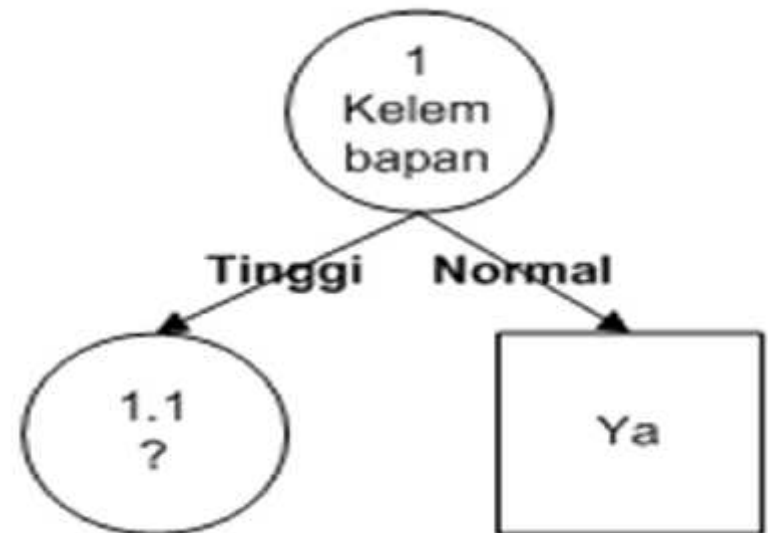
0,86

$$Entropy(Total) - \sum_{i=1}^n \frac{|Cuaca_i|}{|Total|} * Entropy(Cuaca_i)$$

$$0,86 - \left(\left(\frac{4}{14} * 0\right) + \left(\frac{5}{14} * 0,72\right) + \left(\frac{5}{14} * 0,97\right)\right) Gain(Total, Cuaca) = 0,26$$

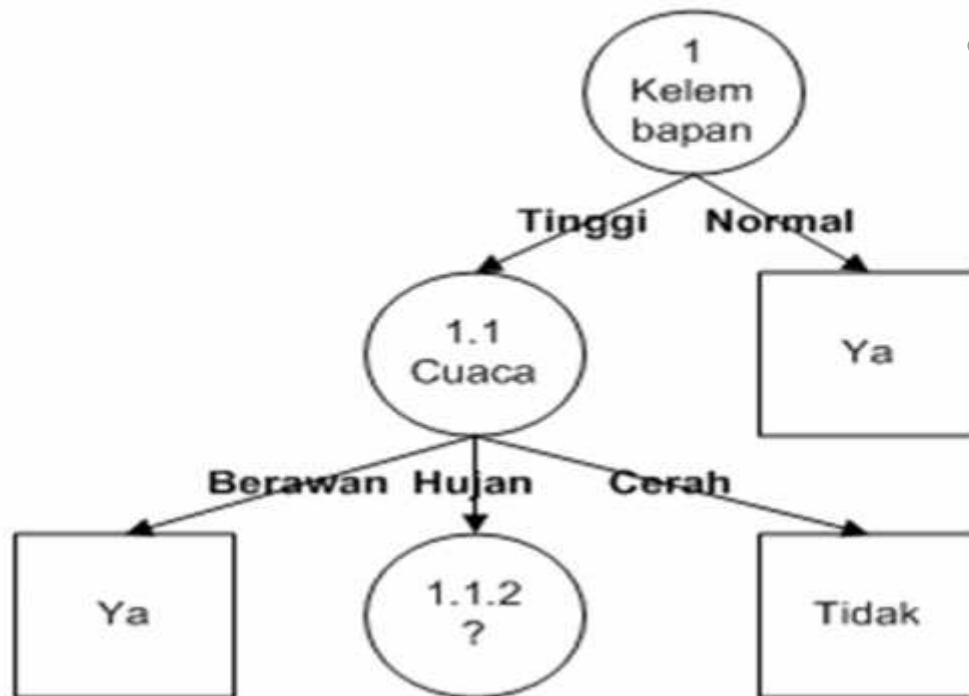
Node			Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1	TOTAL		14	4	10	0.86	
	CUACA						0.26
		BERAWAN	4	0	4		
		HUJAN	5	1	4	0.72	
		CERAH	5	3	2	0.97	
	TEMPERATUR						
		DINGIN					
		PANAS					
		SEDANG					
	KELEMBAPAN						
		TINGGI					
		NORMAL					
	BERANGIN						0.01
		TIDAK	8	2	6	0.81	
		YA	6	4	2	0.92	

NO	CUACA	TEMPERATUR	KELEMBAPAN	BERANGIN	MAIN
1	Cerah	Panas	Tinggi	TIDAK	Tidak
2	Cerah	Panas	Tinggi	YA	Tidak
3	Berawan	Panas	Tinggi	TIDAK	Ya
4	Hujan	Sedang	Tinggi	TIDAK	Ya
5	Hujan	Dingin	Normal	TIDAK	Ya
6	Hujan	Dingin	Normal	YA	Ya
7	Berawan	Dingin	Normal	YA	Ya
8	Cerah	Sedang	Tinggi	TIDAK	Tidak
9	Cerah	Dingin	Normal	TIDAK	Ya
10	Hujan	Sedang	Normal	TIDAK	Ya
11	Cerah	Sedang	Normal	YA	Ya
12	Berawan	Sedang	Tinggi	YA	Ya
13	Berawan	Panas	Normal	TIDAK	Ya
14	Hujan	Sedang	Tinggi	YA	Tidak



Node			Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1	TOTAL		14	4	10	0.86	
	CUACA						0.26
		BERAWAN	4	0	4		
		HUJAN	5	1	4	0.72	
		CERAH	5	3	2	0.97	
	TEMPERATUR						0.18
		DINGIN	4	0	4		
		PANAS	4	2	2	1.00	
		SEDANG	6	2	4	0.92	
	KELEMBAPAN						0.37
		TINGGI	7	4	3	0.99	
		NORMAL	7	0	7		
	BERANGIN						0.01
		TIDAK	8	2	6	0.81	
		YA	6	4	2	0.92	

Node			Σ Kasus (S)	Σ Tidak (S1)	Σ Ya (S2)	Entropy	Gain
1.1	KELEMBAPAN-TINGGI		7	4	3	0,99	
	CUACA						0,7
		BERAWAN	2	0	2	0	
		HUJAN	2	1	1	1	
		CERAH	3	3	0	0	
	TEMPERATUR						0,02
		DINGIN	0	0	0	0	
		PANAS	3	2	1	0,92	
		SEDANG	4	2	2	1	
	BERANGIN						0,02
		TIDAK	4	2	2	1	
		YA	3	2	1	0,92	



- Selesaikan Pohon Tersebut!!!