ALGORITMA C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan.

Tabel 3.1. Keputusan Bermain Tenis

NO	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Dalam kasus yang tertera pada Tabel 3.1, akan dibuat pohon keputusan untuk menentukan main tenis atau tidak dengan melihat keadaan cuaca, temperatur, kelembaban dan keadaan angin.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

a. Pilih atribut sebagai akar
b. Buat cabang untuk masing-masing nilai
c. Bagi kasus dalam cabang
d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk

menghitung gain digunakan rumus seperti tertera dalam Rumus 1 (Craw, S., ---).

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

Dengan :
S : Himpunan kasus
A : Atribut
n : Jumlah partisi atribut A
|Si| : Jumlah kasus pada partisi ke i
|S| : Jumlah kasus dalam S

Sedangkan penhitungan nilai entropy dapat dilihat pada rumus 2 berikut(Craw, S., ---):

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log_2 pi \qquad (2)$$

dengan :
S : Himpunan Kasus
A : Fitur
n : Jumlah partisi S
p_i : Proporsi dari S_i terhadap S

Berikut ini adalah penjelasan lebih rinci mengenai masing-masing langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 untuk menyelesaikan permasalahan pada Tabel 3.1.

a. Menghitung jumlah kasus, jumlah kasus untuk keputusan ***Yes***, jumlah kasus untuk keputusan ***No***, dan Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut ***OUTLOOK***, ***TEMPERATURE***, ***HUMIDITY*** dan ***WINDY***. Setelah itu lakukan penghitungan Gain untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel 3.2.

Tabel 3.2. Perhitungan Node 1

Node			Jml Kasus (S)	Tidak (S_1)	Ya (S_2)	Entropy	Gain
1	TOTAL		14	4	10	0.863120569	
	OUTLOOK						0.258521037
		CLOUDY	4	0	4		
		RAINY	5	1	4	0.721928095	
		SUNNY	5	3	2	0.970950594	
	TEMPERATURE						0.183850925
		COOL	4	0	4	0	
		HOT	4	2	2	1	
		MILD	6	2	4	0.918295834	
	HUMIDITY						0.370506501
		HIGH	7	4	3	0.985228136	
		NORMAL	7	0	7	0	
	WINDY						0.005977711
		FALSE	8	2	6	0.811278124	
		TRUE	6	4	2	0.918295834	

Baris ***TOTAL*** kolom Entropy pada Tabel 3.2 dihitung dengan rumus 2, sebagai berikut:

$$Entropy(Total) = (-\frac{4}{14} * \log_2(\frac{4}{14})) + (-\frac{10}{14} * \log_2(\frac{10}{14}))$$

$$Entropy(Total) = 0.863120569$$

Sementara itu nilai Gain pada baris ***OUTLOOK*** dihitung dengan menggunakan rumus 1, sebagai berikut:

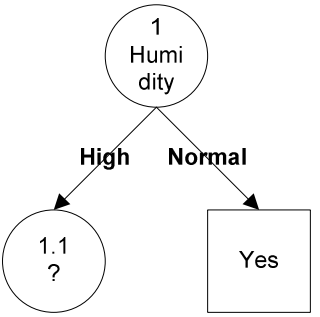
$$Gain(Total, Outlook) = Entropy(Total) - \sum_{i=1}^{n} \frac{|Outlook_i|}{|Total|} * Entropy(Outlook$$

$$Gain(Total, Outlook) = 0.863120569 - ((\frac{4}{14} * 0) + (\frac{5}{14} * 0.723) + (\frac{5}{14} * 0.97)$$

$$Gain(Total, Outlook) = 0.23$$

Dari hasil pada Tabel 3.2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah ***HUMIDITY*** yaitu sebesar 0.37. Dengan demikian ***HUMIDITY*** dapat menjadi node akar. Ada 2 nilai atribut dari ***HUMIDITY*** yaitu ***HIGH*** dan ***NORMAL***. Dari kedua nilai atribut tersebut, nilai atribut ***NORMAL*** sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya ***Yes***, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut ***HIGH*** masih perlu dilakukan perhitungan lagi.

Dari hasil tersebut dapat digambarkan pohon keputusan sementara-nya tampak seperti Gambar 3.1



Gambar 3.1 Pohon Keputusan Hasil Perhitungan Node 1

b. Menghitung jumlah kasus, jumlah kasus untuk keputusan ***Yes***, jumlah kasus untuk keputusan ***No***, dan Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut ***OUTLOOK***, ***TEMPERATURE*** dan ***WINDY*** yang dapat menjadi node akar dari nilai atribut HIGH. Setelah itu lakukan penghitungan Gain untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel 3.3.

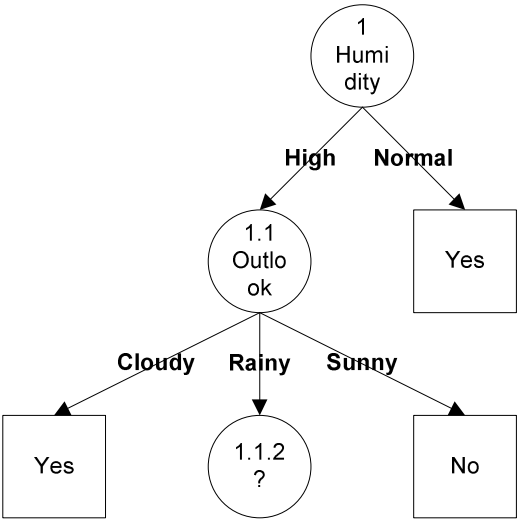
Tabel 3.3. Perhitungan Node 1.1

Node			**Jml Kasus**	**Tidak (S1)**	**Ya (S2)**	**Entropy**	**Gain**

			(S)				
1.1	HUMIDITY-HIGH		7	4	3	0.985228136	
	OUTLOOK						0.69951385
		CLOUDY	2	0	2	0	
		RAINY	2	1	1	1	
		SUNNY	3	3	0	0	
	TEMPERATURE						0.020244207
		COOL	0	0	0	0	
		HOT	3	2	1	0.918295834	
		MILD	4	2	2	1	
	WINDY						0.020244207
		FALSE	4	2	2	1	
		TRUE	3	2	1	0.918295834	

Dari hasil pada Tabel 3.3 dapat diketahui bahwa atribut dengan Gain tertinggi adalah ***OUTLOOK*** yaitu sebesar 0.67. Dengan demikian ***OUTLOOK*** dapat menjadi node cabang dari nilai atribut ***HIGH***. Ada 3 nilai atribut dari ***OUTLOOK*** yaitu ***CLOUDY***, ***RAINY*** dan ***SUNNY***. Dari ketiga nilai atribut tersebut, nilai atribut CLOUDY sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya ***Yes*** dan nilai atribut ***SUNNY*** sudah mengklasifikasikan kasus menjadi satu dengan keputusan ***No***, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut ***RAINY*** masih perlu dilakukan perhitungan lagi.

Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada gambar 3.2 berikut:



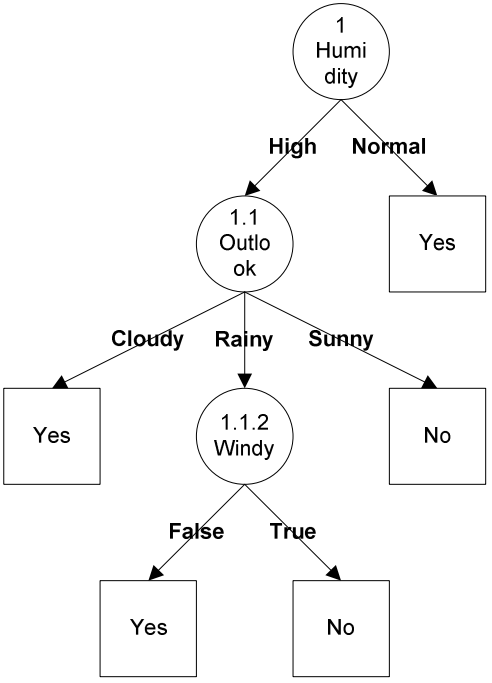
Gambar 3.2. Pohon Keputusan Hasil Perhitungan Node 1.1

c. Menghitung jumlah kasus, jumlah kasus untuk keputusan ***Yes***, jumlah kasus untuk keputusan ***No***, dan Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut ***TEMPERATURE*** dan ***WINDY*** yang dapat menjadi node cabang dari nilai atribut ***RAINY***. Setelah itu lakukan penghitungan Gain untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel 3.4.

   Dari hasil pada tabel 3.4 dapat diketahui bahwa atribut dengan Gain tertinggi adalah ***WINDY*** yaitu sebesar 1. Dengan demikian ***WINDY*** dapat menjadi node cabang dari nilai atribut ***RAINY***. Ada 2 nilai atribut dari ***WINDY*** yaitu ***FALSE*** dan ***TRUE***. Dari kedua nilai atribut tersebut, nilai atribut ***FALSE*** sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya ***Yes*** dan nilai atribut ***TRUE*** sudah mengklasifikasikan kasus menjadi satu dengan keputusan ***No***, sehingga tidak perlu dilakukan perhitungan lebih lanjut untuk nilai atribut ini.

Tabel 3.4. Perhitungan Node 1.1.2

Node			Jml Kasus (S)	Tidak (S1)	Ya (S2)	Entropy	Gain
1.1.2	HUMIDITY-HIGH dan OUTLOOK - RAINY		2	1	1	1	
	TEMPERATURE						0
		COOL	0	0	0	0	
		HOT	0	0	0	0	
		MILD	2	1	1	1	
	WINDY						1
		FALSE	1	0	1	0	
		TRUE	1	1	0	0	



Gambar 3.3. Pohon Keputusan Hasil Perhitungan Node 1.1.2

Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada Gambar 3.3.

Dengan memperhatikan pohon keputusan pada Gambar 3.3, diketahui bahwa semua kasus sudah masuk dalam kelas. Dengan demikian, pohon keputusan pada Gambar 3.3 merupakan pohon keputusan terakhir yang terbentuk.