

NAMA : DIAZ ISLAMI
NIM : A11.2020.13116
KELOMPOK : A11.4411
MATKUL : DATA MINING

JAWABAN – TUGAS 7 DECISION TREE

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
Sunny	Hot	High	No	Don't Play
Sunny	Hot	High	Yes	Don't Play
Cloudy	Hot	High	No	Play
Rainy	Mild	High	No	Play
Rainy	Cool	Normal	No	Play
Rainy	Cool	Normal	Yes	Play
Cloudy	Cool	Normal	Yes	Play
Sunny	Mild	High	No	Don't Play
Sunny	Cool	Normal	No	Play
Rainy	Mild	Normal	No	Play
Sunny	Mild	Normal	Yes	Play
Cloudy	Mild	High	Yes	Play
Cloudy	Hot	Normal	No	Play
Rainy	Mild	High	Yes	Don't Play

- **Hitung Entropy(Total) → Label/Class**

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

$$Entropy(Total) = 0,863120569$$

- **Hitung Entropy(Cloudy) → Feature**

$$Entropy(Cloudy) = \left(-\frac{4}{4} \times \log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right)$$

$$Entropy(Cloudy) = 0$$

- **Hitung Entropy(Rainy) → Feature**

$$Entropy(Rainy) = \left(-\frac{1}{5} \times \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} \times \log_2\left(\frac{4}{5}\right)\right)$$

$$Entropy(Rainy) = 0,72193$$

- **Hitung Entropy(Sunny) → Feature**

$$Entropy(Sunny) = \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} \times \log_2\left(\frac{2}{5}\right)\right)$$

$$Entropy(Sunny) = 0,97095$$

- **Hitung Entropy(Cool) → Feature**

$$Entropy(Cool) = \left(-\frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} \times \log_2\left(\frac{4}{4}\right)\right)$$

$$Entropy(Cool) = 0$$

- **Hitung Entropy(Hot) → Feature**

$$Entropy(Hot) = \left(-\frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right)$$

$$Entropy(Hot) = 1$$

- **Hitung Entropy(Mild) → Feature**

$$Entropy(Mild) = \left(-\frac{2}{6} \times \log_2\left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6} \times \log_2\left(\frac{4}{6}\right)\right)$$

$$Entropy(Mild) = 0,9183$$

- **Hitung Entropy(High) → Feature**

$$Entropy(High) = \left(-\frac{4}{7} \times \log_2\left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} \times \log_2\left(\frac{3}{7}\right)\right)$$

$$Entropy(High) = 0,98523$$

- **Hitung Entropy(Normal) → Feature**

$$Entropy(Normal) = \left(-\frac{0}{7} \times \log_2\left(\frac{0}{7}\right)\right) + \left(-\frac{7}{7} \times \log_2\left(\frac{7}{7}\right)\right)$$

$$Entropy(Normal) = 0$$

- **Hitung Entropy(No) → Feature**

$$Entropy(No) = \left(-\frac{2}{8} \times \log_2\left(\frac{2}{8}\right)\right) + \left(-\frac{6}{8} \times \log_2\left(\frac{6}{8}\right)\right)$$

$$Entropy(No) = 0,81128$$

- **Hitung Entropy(Yes) → Feature**

$$Entropy(Yes) = \left(-\frac{4}{6} \times \log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} \times \log_2\left(\frac{2}{6}\right)\right)$$

$$Entropy(Yes) = 0,9183$$

A. Perhitungan Node 1

- **Hitung Gain(Total ,Outlook)**

$$Gain(Total, Outlook) = 0,863120569 - \left(\left(\frac{4}{14} \times 0 \right) + \left(\frac{5}{14} \times 0,722 \right) + \left(\frac{5}{14} \times 0,97 \right) \right)$$

$$Gain(Total, Outlook) = 0,2585$$

- **Hitung Gain(Total ,Temp)**

$$Gain(Total, Temp) = 0,863120569 - \left(\left(\frac{4}{14} \times 0 \right) + \left(\frac{4}{14} \times 1 \right) + \left(\frac{6}{14} \times 0,9183 \right) \right)$$

$$Gain(Total, Temp) = 0,1838509$$

- **Hitung Gain(Total ,Humidity)**

$$Gain(Total, Humidity) = 0,863120569 - \left(\left(\frac{7}{14} \times 0,985 \right) + \left(\frac{7}{14} \times 0 \right) \right)$$

$$Gain(Total, Humidity) = 0,3705065$$

- **Hitung Gain(Total ,Windy)**

$$Gain(Total, Windy) = 0,863120569 - \left(\left(\frac{8}{14} \times 0,811 \right) + \left(\frac{6}{14} \times 0,9183 \right) \right)$$

$$Gain(Total, Windy) = 0,0059777$$

- **Tabel Perhitungan Node 1**

		Jml kasus(S)	Don't(S1)	Play(S2)	Entropy	Gain
total		14	4	10	0,86312	
outlook						0,258521
	cloudy	4	0	4	0	
	rainy	5	1	4	0,72193	
	sunny	5	3	2	0,97095	
temp						
	col	4	0	4	0	
	hot	4	2	2	1	
	mild	6	2	4	0,9183	
humidity						0,3705065
	high	7	4	3	0,98523	
	normal	7	0	7	0	
windy						0,0059777
	no	8	2	6	0,81128	
	yes	6	4	2	0,9183	

Kesimpulan perhitungan node 1: Karena jumlah gain(total, humidity) lebih besar daripada yang lain, maka humidity menjadi root dalam decision tree. Pada humidity, terdapat 2

features yaitu “high” dan “normal” yang mana pada feature “normal” mengklasifikasikan kasus menjadi 1, yaitu keputusan “Play” sehingga tidak perlu dilakukan perhitungan lebih lanjut. Tetapi, untuk nilai “high” masih perlu dilakukan perhitungan lagi untuk humidity bernilai “high” dengan kasus klasifikasi “Play”. **Karena konsep perhitungan dari entropi dan gain terhadap entropi sama dengan rumus perhitungan sebelumnya, maka pada perhitungan entropi dan gain terhadap entropi tidak akan disajikan secara langsung, namun disajikan dalam tabel hasil dari perhitungan tersebut.**

- **Tabel Perhitungan Node 1.1**

		Jml kasus(S)	Don't(S1)	Play(S2)	Entropy	Gain
humidity		7	4	3	0,98522814	
high						
outlook						0,69951385
	cloudy	2	0	2	0	
	rainy	2	1	1	1	
	sunny	3	3	0	0	
temp						0,02024421
	col	0	0	0	0	
	hot	3	2	1	0,91829583	
	mild	4	2	2	1	
windy						0,02024421
	no	4	2	2	1	
	yes	3	2	1	0,91829583	

Kesimpulan perhitungan node 1.1: Karena jumlah gain(humidity high,outlook) lebih besar daripada yang lain, maka humidity menjadi cabang dari root dalam decision tree. Pada outlook, terdapat 3 features yaitu “cloudy”, “rainy” dan “sunny” yang mana pada feature “cloudy” mengklasifikasikan kasus menjadi 1, yaitu keputusan “Play” sehingga tidak perlu dilakukan perhitungan lebih lanjut; feature “sunny” mengklasifikasikan kasus menjadi 1, yaitu keputusan “Don’t play” sehingga tidak perlu dilakukan perhitungan lebih lanjut. Tetapi, untuk nilai “rainy” masih perlu dilakukan perhitungan lagi untuk outlook bernilai “rainy” dengan kasus klasifikasi “Play” dan “Don’t play”.

- **Tabel Perhitungan Node 1.1.2**

		Jml kasus(S)	Don't(S1)	Play(S2)	Entropy	Gain
humidity		2	1	1	1	
high and outlook						
rainy						
temp						0
	col	0	0	0	0	
	hot	0	0	0	0	
	mild	2	1	1	1	
windy						1
	no	1	0	1	0	
	yes	1	1	0	0	

Kesimpulan perhitungan node 1.1.2: Karena jumlah gain(humidity high & outlook rainy,windy) lebih besar daripada yang lain, maka humidity menjadi cabang dari outlook dengan nilai “rainy”

dalam decision tree. Pada windy, terdapat 2 features yaitu “no” dan “yes” yang mana pada kedua feature tersebut mengklasifikasikan kasus menjadi 1, yaitu untuk “no” ialah keputusan “Play” serta untuk “yes” ialah keputusan “Don’t Play” sehingga tidak perlu dilakukan perhitungan lebih lanjut. Sehingga dari kesimpulan yang telah diperhitungkan, dapat terbentuk decision tree dari kasus keputusan bermain tenis sebagai berikut:

