

Nama : Margareta Valencia

NIM : A11.2022.14704

Kelompok : A11.4509

SOAL KUIS 1 PERTEMUAN 7

Hitung Entropy dan Gain serta tentukan pohon keputusan yang terbentuk dari contoh kasus keputusan bermain tenis dibawah ini :

NO	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	No	Don't Play
2	Sunny	Hot	High	Yes	Don't Play
3	Cloudy	Hot	High	No	Play
4	Rainy	Mild	High	No	Play
5	Rainy	Cool	Normal	No	Play
6	Rainy	Cool	Normal	Yes	Play
7	Cloudy	Cool	Normal	Yes	Play
8	Sunny	Mild	High	No	Don't Play
9	Sunny	Cool	Normal	No	Play
10	Rainy	Mild	Normal	No	Play
11	Sunny	Mild	Normal	Yes	Play
12	Cloudy	Mild	High	Yes	Play
13	Cloudy	Hot	Normal	No	Play
14	Rainy	Mild	High	Yes	Don't Play

JAWAB

Rumus

1. Entropy (S) = $\sum_{i=1}^n -p_i * \log_2 p_i$
2. Gain (S, A) = Entropy(S) - $\sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$

Langkah 1

Menghitung jumlah kasus, jumlah kasus untuk keputusan **Yes** dan **No**, serta Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut **OUTLOOK**, **TEMPERATURE**, **HUMIDITY**, dan **WINDY**.

Detail perhitungan :

$$\text{Entropy}(\text{Total}) = \left(-\frac{4}{14} \cdot \log_2\left(\frac{4}{14}\right)\right) + \left(-\frac{10}{14} \cdot \log_2\left(\frac{10}{14}\right)\right) = 0,863120569$$

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

$$= 0,863120569 - \left(\left(\frac{4}{14} * 0\right) + \left(\frac{5}{14} * 0,722\right) + \left(\frac{5}{14} * 0,97\right)\right) = 0,258521037$$

$$\text{a. Entropy (cloudy)} = \left(-\frac{0}{4} \cdot \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} \cdot \log_2\left(\frac{4}{4}\right)\right) = 0$$

$$\text{b. Entropy (rainy)} = \left(-\frac{1}{5} \cdot \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} \cdot \log_2\left(\frac{4}{5}\right)\right) = 0,721928095$$

$$\text{c. Entropy (sunny)} = \left(-\frac{3}{5} \cdot \log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} \cdot \log_2\left(\frac{2}{5}\right)\right) = 0,970950594$$

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

$$= 0,863120569 - \left(\left(\frac{4}{14} * 0\right) + \left(\frac{4}{14} * 1\right) + \left(\frac{6}{14} * 0,92\right)\right) = 0,183850925$$

$$\text{a. Entropy (cool)} = \left(-\frac{0}{4} \cdot \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} \cdot \log_2\left(\frac{4}{4}\right)\right) = 0$$

$$\text{b. Entropy (hot)} = \left(-\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right)\right) = 1$$

$$\text{c. Entropy (mild)} = \left(-\frac{2}{6} \cdot \log_2\left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6} \cdot \log_2\left(\frac{4}{6}\right)\right) = 0,918295834$$

$$\text{Gain}(\text{Total, Humidity}) = \text{Entropy}(\text{Total}) - \sum_{i=1}^n \frac{|\text{Humidity}_i|}{|\text{Total}|} * \text{Entropy}(\text{Hum}_i)$$

$$= 0,863120569 - \left(\left(\frac{7}{14} * 0,918\right) + \left(\frac{7}{14} * 0\right)\right) = 0,370506501$$

$$\text{a. Entropy (high)} = \left(-\frac{4}{7} \cdot \log_2\left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} \cdot \log_2\left(\frac{3}{7}\right)\right) = 0,985228136$$

$$b. \text{ Entropy (normal)} = \left(-\frac{0}{7} \cdot \log_2 \left(\frac{0}{7}\right)\right) + \left(-\frac{7}{7} \cdot \log_2 \left(\frac{7}{7}\right)\right) = 0$$

$$\text{Gain (Total, Windy)} = \text{Entropy(Total)} - \sum_{i=1}^n \frac{|Windy_i|}{|Total|} * \text{Entropy(Windy}_i)$$

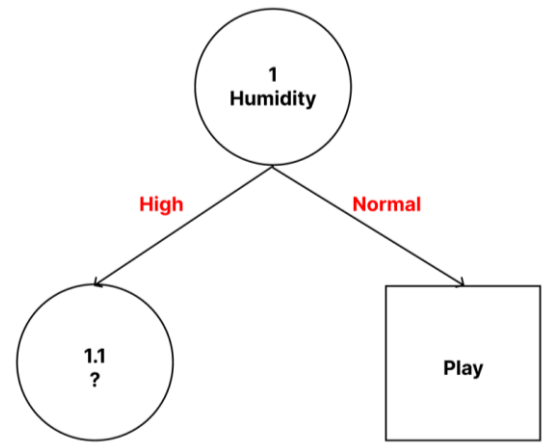
$$= 0.863120569 - \left(\left(\frac{8}{14} * 0.811\right) + \left(\frac{6}{14} * 0.918\right)\right) = 0.005977713$$

$$a. \text{ Entropy (no)} = \left(-\frac{2}{8} \cdot \log_2 \left(\frac{2}{8}\right)\right) + \left(-\frac{6}{8} \cdot \log_2 \left(\frac{6}{8}\right)\right) = 0,811278124$$

$$b. \text{ Entropy (yes)} = \left(-\frac{4}{6} \cdot \log_2 \left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} \cdot \log_2 \left(\frac{2}{6}\right)\right) = 0,91829583$$

Node 1		jml kasus(S)	No (S1)	Yes (S2)	Entropy	Gain
total		14	4	10	0,863120569	
outlook						0,258521037
	cloudy	4	0	4	0	
	rainy	5	1	4	0,721928095	
	sunny	5	3	2	0,970950594	
temp						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						0,005977713
	No	8	2	6	0,811278124	
	Yes	6	4	2	0,91829583	

Dari hasil perhitungan diketahui bahwa atribut dengan gain tertinggi adalah **HUMIDITY** dengan nilai **0,370506501**. Oleh karena itu, HUMIDITY menjadi node akar dari pohon keputusan. HUMIDITY memiliki dua nilai atribut, yaitu **HIGH** dan **NORMAL**. Untuk cabang dengan atribut HUMIDITY NORMAL, dapat langsung diklasifikasikan sebagai leaf node dengan keputusan "Yes" (Play), karena seluruh instance dengan nilai HUMIDITY NORMAL konsisten menghasilkan keputusan yang sama. Sebaliknya, cabang dengan atribut HUMIDITY HIGH memerlukan analisis lebih lanjut. Hal ini menunjukkan bahwa data dengan nilai HUMIDITY HIGH memiliki variasi yang perlu diteliti lebih dalam sebelum dapat ditentukan keputusannya secara pasti.



Langkah 2

Menghitung jumlah kasus, jumlah kasus untuk keputusan **Yes**, jumlah kasus untuk keputusan **No**. Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut **OUTLOOK, TEMPERATURE** dan **WINDY**, yang dapat menjadi node akar dari nilai atribut **HIGH**.

Detail perhitungan :

$$\text{Entropy}(S) = \left(-\frac{3}{7} \cdot \log_2 \left(\frac{3}{7}\right)\right) + \left(-\frac{4}{7} \cdot \log_2 \left(\frac{4}{7}\right)\right) = 0,98522814$$

$$\begin{aligned} \text{Gain}(S, \text{Outlook}) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|\text{Outlook}_i|}{|S|} * \text{Entropy}(\text{Outlook}_i) \\ &= 0.98522814 - \left(\left(\frac{2}{7} * 0\right) + \left(\frac{2}{7} * 1\right) + \left(\frac{3}{7} * 0\right)\right) = 0.69951385 \end{aligned}$$

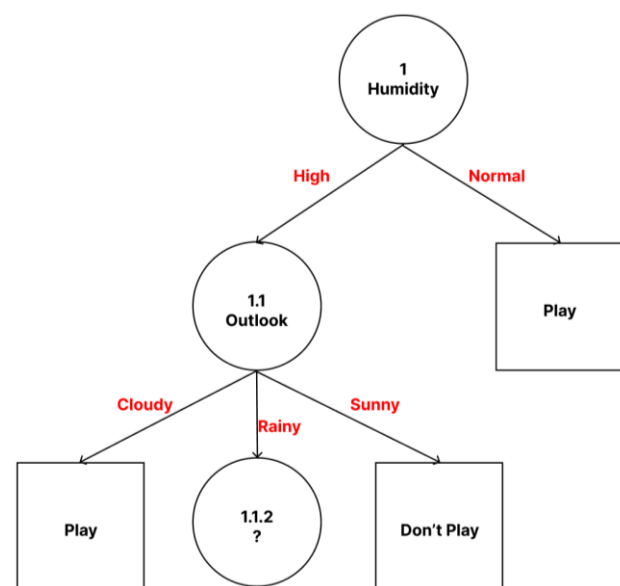
$$\begin{aligned} \text{Gain}(S, \text{Temp}) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|\text{Temp}_i|}{|S|} * \text{Entropy}(\text{Temp}_i) \\ &= 0.98522814 - \left(\left(\frac{0}{7} * 0\right) + \left(\frac{3}{7} * 0.9182\right) + \left(\frac{4}{7} * 1\right)\right) = 0.02024421 \end{aligned}$$

$$\begin{aligned} \text{Gain}(S, \text{Windy}) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|\text{Windy}_i|}{|S|} * \text{Entropy}(\text{Windy}_i) \\ &= 0.98522814 - \left(\left(\frac{4}{7} * 1\right) + \left(\frac{3}{7} * 0.9182\right)\right) = 0.02024421 \end{aligned}$$

Node 1.1		jml kasus(S)	No (S1)	Yes (S2)	Entropy	Gain
humidity (high)		7	4	3	0,98522814	
outlook						0,69951385
	cloudy	2	0	2	0	
	rainy	2	1	1	1	
	sunny	3	3	0	0	
temp						0,02024421
	cool	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	

OUTLOOK dipilih sebagai node cabang dari nilai HUMIDITY = High karena Gain tertinggi sebesar **0.69951385**. OUTLOOK memiliki tiga nilai:

- CLOUDY : Semua kasus di cabang ini diklasifikasikan sebagai Play.
- SUNNY : Semua kasus di cabang ini diklasifikasikan sebagai Don't Play.
- RAINY : Masih ada variasi antara Play dan Don't Play, sehingga perlu dilakukan perhitungan lebih lanjut untuk atribut lain agar bisa mengklasifikasikan kasus di cabang ini.



Langkah 3

Menghitung jumlah kasus, jumlah kasus untuk keputusan Yes, jumlah kasus untuk keputusan No. Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut TEMPERATURE dan WINDY, yang dapat menjadi node cabang dari nilai atribut RAINY.

Detail perhitungannya :

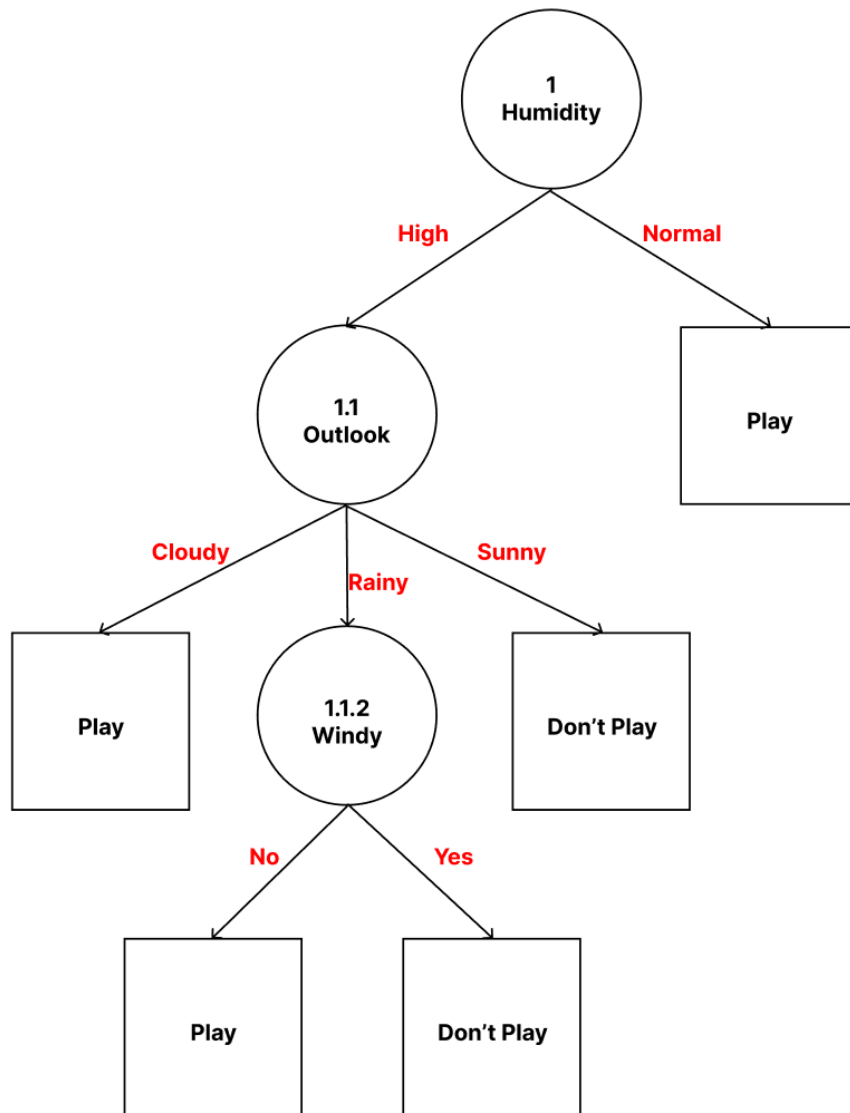
$$\text{Entropy}(S) = \left(-\frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$\begin{aligned} \text{Gain}(S, \text{Temp}) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|\text{Temp}_i|}{|S|} * \text{Entropy}(\text{Temp}_i) \\ &= 1 - \left(\left(\frac{0}{2} * 0\right) + \left(\frac{3}{0} * 0\right) + \left(\frac{2}{2} * 1\right)\right) = 0 \end{aligned}$$

$$\begin{aligned} \text{Gain}(S, \text{Windy}) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|\text{Windy}_i|}{|S|} * \text{Entropy}(\text{Windy}_i) \\ &= 1 - \left(\left(\frac{1}{2} * 0\right) + \left(\frac{1}{2} * 0\right)\right) = 1 \end{aligned}$$

Node 1.1.2		jml kasus(S)	No (S1)	Yes (S2)	Entropy	Gain
humidity high and outlook rainy		2	1	1	1	
temp						0
	cool	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	

Atribut **WINDY** dengan Gain tertinggi (1) menjadi node untuk OUTLOOK = RAINY. Jika WINDY = No, semua kasus adalah Play, dan jika WINDY = Yes, maka semua kasus adalah Don't Play. Karena kedua nilai sudah mengklasifikasikan kasus, maka tidak perlu perhitungan lebih lanjut.



SOAL KUIS 2 PERTEMUAN 7

Link : <https://github.com/margaretavalen/Data-Mining/tree/main/Tugas7>