

LITS DATA MINING

NIM 17129000608

NAMA MUHAMMAD ILHAM STAFI'I

PAJDI TEKNIK INFORMATIKA K2

SEMESTER 6

1) Informasi, Algoritma, & Data Mining

1. Interpretasi Data dan mencari nilai entropi dari data:

$$= \text{Entropy}(S) = -\sum p_i \log_2 p_i$$

1. Buat cabang untuk tiap atributnya

2. Meng pilih untuk cabang terbaik dengan cara mencari nilai entropi masing-masing kelas yang ada

Contoh

NIM	Nilai IPK	Angka Ujian	Informasi	Apakah Baik
1	sedang	2.100-2.500	ada	ya
2	rendah	2.500-3.000	tidak ada	tidak
3	tinggi	3.000-3.500	ada	ya
4	rendah	2.100-2.500	tidak ada	tidak
5	tinggi	2.600-3.000	ada	ya
6	tinggi	2.100-2.500	tidak ada	ya
7	sedang	2.100-2.500	ada	ya
8	rendah	2.100-2.500	ada	ya
9	rendah	2.500-3.000	ada	ya
10	tinggi	2.500-3.000	tidak ada	tidak
11	sedang	3.000-3.500	tidak ada	ya

$$2. \text{Total Data (S)} = 11, \text{ Ya (Y)} = 8, \text{ No (N)} = 3, \text{ Entropy}(S) = -\sum p_i \log_2 p_i$$

3. Pilih atribut sebagai atribut akar (Nilai IPK)

Nilai IPK	Jumlah Data (S)	Ya (Y)	No (N)	Entropy (S)
rendah	3	2	1	= 0,918291634
sedang	4	3	1	
tinggi	4	3	1	

$$= \text{Entropy}(\text{Nilai IPK, rendah}) = (-\frac{2}{3} \log_2 \frac{2}{3}) - (\frac{1}{3} \log_2 \frac{1}{3})$$

$$= \text{Entropy}(\text{Nilai IPK, sedang}) = (-\frac{3}{4} \log_2 \frac{3}{4}) - (\frac{1}{4} \log_2 \frac{1}{4})$$

* Untuk bisa belajar coding adalah penting, maka cobalah
 memilih bahasa yang akan dipelajari



3	Preferensi	Nilai p_k	Penghasilan x_k	Dipilih $Bernilai$
	tidak ada	bedang	2.300.000	tidak
	tidak ada	pendek	2.100.000	tidak
	tidak ada	Tinggi	2.200.000	Ya
	tidak ada	Tinggi	2.100.000	tidak
	tidak ada	bedang	2.200.000	Ya

Total kasus 5 2 3 Entropy total dari seluruh masalah
 = Entropy total

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

$$= 0.92092599$$

* Berikut cabang untuk top-down:

Node	Simbol	Label Kasus	Ya	Tidak	Entropy	Gain
	Total	5	2	3	0.92092599	
1	Nilai p_k Rendah	1	0	1	0	0.17095039
	bedang	1	1	1	1	
	Tinggi	2	1	1	1	
	2.1000.000	3	2	1	0.918295834	0.9198299
	2.3000.000	2	0	2	0	
Penghasilan x_k						

$$\text{Entropy (Nilai } p_k \text{ Rendah)} = \left(-\frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right) + \left(-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right)$$

$$= 0$$

$$\text{Entropy (Nilai } p_k \text{ bedang)} = \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right)$$

$$= 1$$

$$\text{Entropy (Nilai } p_k \text{ Tinggi)} = \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right) + \left(-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right)$$

$$= 1$$

$$\text{Entropy (Gulap - Bayu)} = (-\frac{3}{10} \cdot \log_2(\frac{3}{10})) + (-\frac{7}{10} \cdot \log_2(\frac{7}{10}))$$

$$= 0,81178110$$

* Info Gain sebagai akar pohon pengklasifikasi orang tua

Penghasilan orang tua	Jumlah orang tua	Ya	Tidak
< 2000.000	3	3	0
2.000.000	5	4	1
7.000.000	3	1	2

$$\text{Entropy (Penghasilan orang tua, <2000.000)} = (-\frac{3}{3} \cdot \log_2(\frac{3}{3})) + (-\frac{0}{3} \cdot \log_2(\frac{0}{3})) = 0$$

$$\text{Entropy (Penghasilan orang tua, 2.000.000)} = (-\frac{4}{5} \cdot \log_2(\frac{4}{5})) + (-\frac{1}{5} \cdot \log_2(\frac{1}{5}))$$

$$= 0,721928095$$

$$\text{Entropy (Penghasilan orang tua, 7.000.000)} = (-\frac{1}{3} \cdot \log_2(\frac{1}{3})) + (-\frac{2}{3} \cdot \log_2(\frac{2}{3}))$$

* Info absolut sebagai akar pohon = 0,918295834

Waktu	Jumlah orang tua	Ya	Tidak
Arida	6	6	0
Tidak Arida	5	2	3

$$\text{Entropy (Arida, Arida)} = (-\frac{6}{6} \cdot \log_2(\frac{6}{6})) + (-\frac{0}{6} \cdot \log_2(\frac{0}{6}))$$

$$= 0$$

$$\text{Entropy (Arida, Tidak Arida)} = (-\frac{2}{5} \cdot \log_2(\frac{2}{5})) + (-\frac{3}{5} \cdot \log_2(\frac{3}{5}))$$

$$= 0,920925804$$

Menghitung Nilai Entropi untuk masing-masing atribut

Indeks	Atribut	Jumlah	Ya	Tidak	Entropy
	Total	11	7	4	0,845350937
1	Nilai PPK				
	Arida	3	3	0	0,918295834
	Sedang	4	3	1	
	Tinggi	4	1	3	

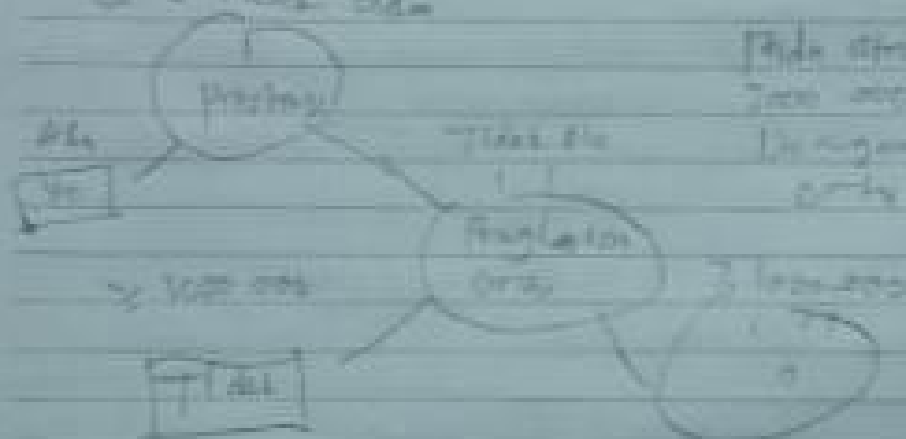
$$G_{\text{orang}}(\text{penghasilan oris} = 200000) = (2/3) \cdot \log_2(1/3) + (1/3) \cdot \log_2(2/3) \\ = 0,918295834$$

$$G_{\text{orang}}(\text{penghasilan oris} > 500000) = (1/2) \cdot \log_2(1/2) + (1/2) \cdot \log_2(1/2) \\ = 1$$

$$G_{\text{orang}}(\text{tidak, tidak ada}) = 0,999999999 - ((1/3) \cdot 0) + ((2/3) \cdot 0,918295834) \\ = 0,999999999$$

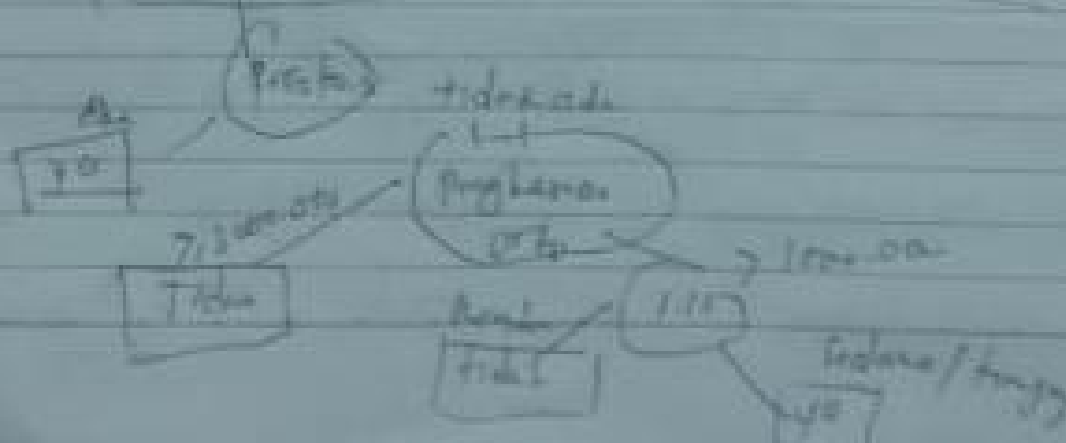
$$G_{\text{orang}}(\text{tidak, penghasilan oris}) = 0,999999999 - ((1/3) \cdot 0,918295834) + ((2/3) \cdot 1) = 0,999999999$$

Maka, Gini terbesar adalah untuk penghasilan oris 100000
 Akibat penghasilan oris menjadi 100000 orang dan tidak ada
 = tidak ada



Maka untuk penghasilan oris > 700000
 dan ada maka akan ada
 dengan penghasilan
 oris = 700000
 menjadi dua / ya /

No.	Metode	Jumlah		
		Ya	Tidak	
1.1.1	Penghasilan oris < 700000	3	2	
	Alasan: Rendah	1	0	1
	Alasan: Tinggi	1	1	0
	Alasan: Tinggi	1	1	0



1) Handwritten Data Testing Handwritten Number Data

Time	Handwritten Data	Handwritten Data	Handwritten Data	Handwritten Data
1	2.5	2.5	2.5	2.5
2	3.5	3.5	3.5	3.5
3	4.5	4.5	4.5	4.5
4	5.5	5.5	5.5	5.5