Nama : Andika Jati F
NIm : 3022110001
Mata kuliah : Penambangan Data

1. Hitunglah secara manual information gain dari data disamping
   - Information gain (class; online course)
   - Information gain (class; Education background)
   - Information gain (class; working status)

online course	Education Background	working status	class
Y	maths	NW	Pass
N	maths	W	Fail
Y	maths	W	Fail
Y	CS	NW	pass
N	other	W	Fail
Y	other	NW	Fail
Y	maths	NW	pass
Y	CS	NW	pass
N	maths	W	pass
N	CS	W	pass
Y	CS	W	pass
N	maths	NW	pass
Y	other	W	Fail
N	other	NW	Fail
N	maths	W	Fail

Testing

• Result	
Pass	Fail
8	4

$$E(\text{Result}) = -\frac{8}{12} \times \log_2\left(\frac{8}{12}\right) - \frac{4}{12} \times \log_2 \frac{4}{12}$$

$$= -0,667 \times \log_2(0,667) - 0,333 \times \log_2(0,333)$$

$$= -0,667 \times (-0,584) - 0,333 \times (-1,586)$$

$$= 0,917$$

Result Online course

	pass	fail	sum
Y	5	2	7
N	3	2	5

sum = 12

$$E(Y) = -\frac{5}{7} \times \log_2\left(\frac{5}{7}\right) - \frac{2}{7} \times \log_2\left(\frac{2}{7}\right)$$
$$= -0,714 \times \log_2(0,714) - 0,285 \times \log_2(0,285)$$
$$= -0,714 \times (-0,486) - 0,285 \times (-1,811)$$
$$= 0,863$$

$$E(N) = -\frac{3}{5} \times \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \times \log_2\left(\frac{2}{5}\right)$$
$$= -0,6 \times \log_2(0,6) - 0,4 \times \log_2(0,4)$$
$$= -0,6 \times (-0,737) - 0,4 \times (-1,322)$$
$$= 0,971$$

$$E(\text{Result: online course}) = P(Y) \times E(Y) + P(N) \times E(N)$$
$$= 7/12 \times 0,863 + 5/12 \times 0,971$$
$$= 0,908$$

Result Education background

	pass	fail	sum
maths	4	2	6
cs	4	0	4
other	0	2	2

sum = 12

$E(\text{maths}) = -4/6 \times \log_2(4/6) - 2/6 \times \log_2(2/6)$

$= -0,667 \times \log_2(0,667) - 0,333 \times \log_2(0,333)$

$= -0,667 \times (-0,584) - 0,333 \times (-1,586)$

$= 0,917$

$E(CS) = -4/4 \times \log_2(4/4) - 0/4 \times \log_2(0/4)$

$= 0 \times \log_2(1) - 0 \times \log_2(0)$

$= -1 \times 0 - 0$

$= 0$

$E(\text{other}) = -0/2 \times \log_2(0/2) - 2/2 \times \log_2(2/2)$

$= 0 \times \log_2(0) - 1 \times \log_2(1)$

$= 0 - 1 \times 0$

$= 0$

$E(R : \text{Education background}) = P(\text{maths}) \times E(\text{maths}) + P(CS) + P(\text{other}) \times E(\text{other})$

$= 6/12 \times 0,917 + 4/12 \times 0 + 2/12 \times 0$

$= 0,5 \times 0,917 + 0 + 0$

$= 0,458$

Result working status

	Pass	fail	sum
W	4	3	7
NW	5	0	5
		Sum = 12	

$E(vi) = -4/7 \times \log_2(4/7) - 3/7 \times \log_2(3/7)$

$= -0,571 \times \log_2(0,571) - 0,428 \times \log_2(0,428)$

$= -0,571 \times (-0,808) - 0,428 \times (-1,224)$

$= 0,985$

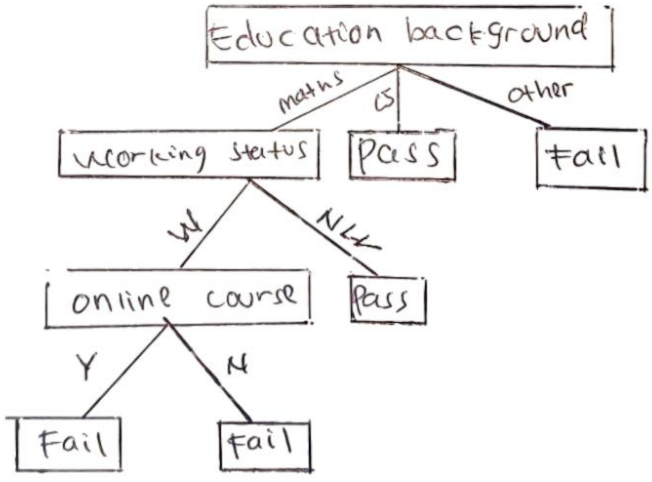
$$E(NW) = -\frac{5}{5} \times \log_2\left(\frac{5}{5}\right) - \frac{0}{5} \times \log_2\left(\frac{0}{5}\right)$$
$$= -1 \times \log_2(1) - 0 \times \log_2(0)$$
$$= -1 \times 0 - 0$$
$$= 0$$

$$E(R: \text{working status}) = P(W) \times E(W) + P(NW) \times E(NW)$$
$$= 7/12 \times 0{,}985 + 5/12 \times 0$$
$$= 0{,}574$$

E (R: online course) $= 0{,}908$
E (R: Education background) $= 0{,}458$
E (R: working status) $= 0{,}574$

E (Result) $= 0{,}917$

Information gain (R: O-C) $= 0{,}917 - 0{,}908 = 0{,}006$
Information gain (R: E-B) $= 0{,}917 - 0{,}458 = 0{,}459$
Information gain (R: W-S) $= 0{,}917 - 0{,}574 = 0{,}343$



R	O_C	E_B	W_S
Fail	Y	other	Fail
Fail	N	other	Fail
Fail	N	maths	Fail

2. Diketahui sebuah problem klasifikasi 3 kelas memiliki confusion matrix sebagai berikut

- Hitunglah akurasi keseluruhan classifier tersebut
- berapa presisi class C_2

Kelas Aktual	C_1	C_2	C_3
C_1	110	8	7
C_2	16	130	10
C_3	26	5	120

$$\text{Akurasi} = \frac{\text{Jumlah prediksi yang benar}}{\text{Total Jumlah prediksi}}$$

$$\text{\# Akurasi} = \frac{360}{432} = 0,8333$$
$$= 83,33\%$$

$$\text{Presisi } (C_2) = \frac{\text{True positive } C_2}{\text{True positive } C_2 + \text{False positive } C_2}$$

False positive $C_2 = 8 + 5 = 13$

$$\text{Presisi } (C_2) = \frac{130}{130 + 13} = \frac{130}{143} = 0,9091$$
$$= 90,91\%$$