(*) Construct Decision tree for the following transactions using ID3

TID	Refund	Martial Status	Taxable income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	NO
8	No	Single	85K	Yes.
9	No	Married	75K	No
10	No	Single	90K	Yes

Let Consider attribute Taxableincome

Sort The Continious Values of attribute Taxableincome

	Cheat	
60k	→ No	
70k	→ No	
75k	→ No	} Change in the Target Concept Value
85k	→ Yes	
90k	→ Yes	
95k	→ Yes	} Change in the Target Concept Value
100k	→ No	
120k	→ No	
125k	→ No	
220k	→ No	

In this,
We should picks a Threshold, c, That produces the greatest Informating gain.

∴ Identifying adjacent examples That Differ in their target classificating.

From this, ~~Coolm~~ Thresolds are

$$C_1 = \frac{75k + 85k}{2} = 80$$

$$C_2 = \frac{95k + 100k}{2} = 97.5$$

Now,

We have to Calculate Information Gain on attribute Taxable Income

Let there Threshold

$C_1 = 80$

$C_2 = 97.5$

Let's Consider C_1

Before That, Calculate Entropy(D).

$$\text{Entropy}(D) = -\sum_{i=1}^{2} P_i \log_2 P_i$$

(Info(D))

$$= -\left[\left(\frac{3}{10}\right)\log_2\left(\frac{3}{10}\right) + \left(\frac{7}{10}\right)\log_2\left(\frac{7}{10}\right)\right]$$

$$= 0.88 \text{ bits}$$

Info (D)

Taxable income wrt $C_1 > 80$k

$$= \sum_{j=1}^{2} \frac{|D_j|}{|D|} \text{Info}(D_j)$$

$$= \left(\frac{7}{10}\right) \text{Info}\left(D_{C_1 > 80k}\right) + \left(\frac{3}{10}\right) \text{Info}\left(D_{C_1 \leq 80}\right)$$

~~$\left(\frac{7}{10}\right)\left[-\left(\frac{3}{7}\right)\log_2\left(\frac{3}{7}\right)\right]$~~

$$= \frac{7}{10}\left[-\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7}\right]$$

$$+ \frac{3}{10}\left[-\frac{0}{3}\log_2\left(\frac{0}{3}\right) - \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right]$$ (0)

$$= 0.68$$

Gain (Taxable income wrt $C_1 > 80$k) = Info(D) − Info(D) Taxable income wrt $C_1 > 80$k

$$= 0.88 - 0.68$$

$$= 0.20$$

$\widehat{Info(D)}$

Taxable income

w.r.t $C_2 > 97.5$

$$= \phi \sum_{j=1}^{2} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \left(\frac{4}{10}\right)\left[-\left(\frac{0}{4}\right)\log_2\left(\frac{0}{4}\right) - \left(\frac{4}{4}\right)\log_2\left(\frac{4}{4}\right)\right] \quad (-\,0)$$

$$+$$

$$\left(\frac{6}{10}\right)\left[-\left(\frac{3}{6}\right)\log_2\left(\frac{3}{6}\right) - \left(\frac{3}{6}\right)\log_2\left(\frac{3}{6}\right)\right]$$

$$= \underline{0.6}$$

Gain (Taxable income)

wrt $C_2 > 97.5$

$$= 0.88 - 0.6$$

$$= \underline{0.28}$$

Next Consider attribut Martial Status

$$\underset{\text{Marital Status}}{Info(D)} = \sum_{j=1}^{3} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{single}|}{|D|} Info(D_{single}) + \frac{|D_{married}|}{|D|} Info(D_{married}) + \frac{|D_{Divorced}|}{|D|} Info(D_{Divorced})$$

$$= \left(\frac{4}{10}\right)\left[-\left(\frac{2}{4}\right)\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2(2/4)\right]$$

$$+ \left(\frac{4}{10}\right)\left[-\left(\frac{0}{4}\right)\log_2\left(\frac{0}{4}\right) - \frac{4}{4}\log_2\left(\frac{4}{4}\right)\right] \quad (0)$$

$$+ \left(\frac{2}{10}\right)\left[-\left(\frac{1}{2}\right)\log_2\left(\frac{1}{2}\right) - 1/2\log_2(1/2)\right]$$

$$= 0.4 + 0.2$$

$$= \underline{0.6}$$

Gain (Marital Status)

$$= 0.88 - 0.6$$

$$= \underline{0.28}$$

Next Consider attribute Refund

$$Info(D)$$

$$Info_{Refund}(D) = \sum_{j=1}^{2} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{yes}|}{|D|} Info(D_{yes}) + \frac{|D_{No}|}{|D|} Info(D_{No})$$

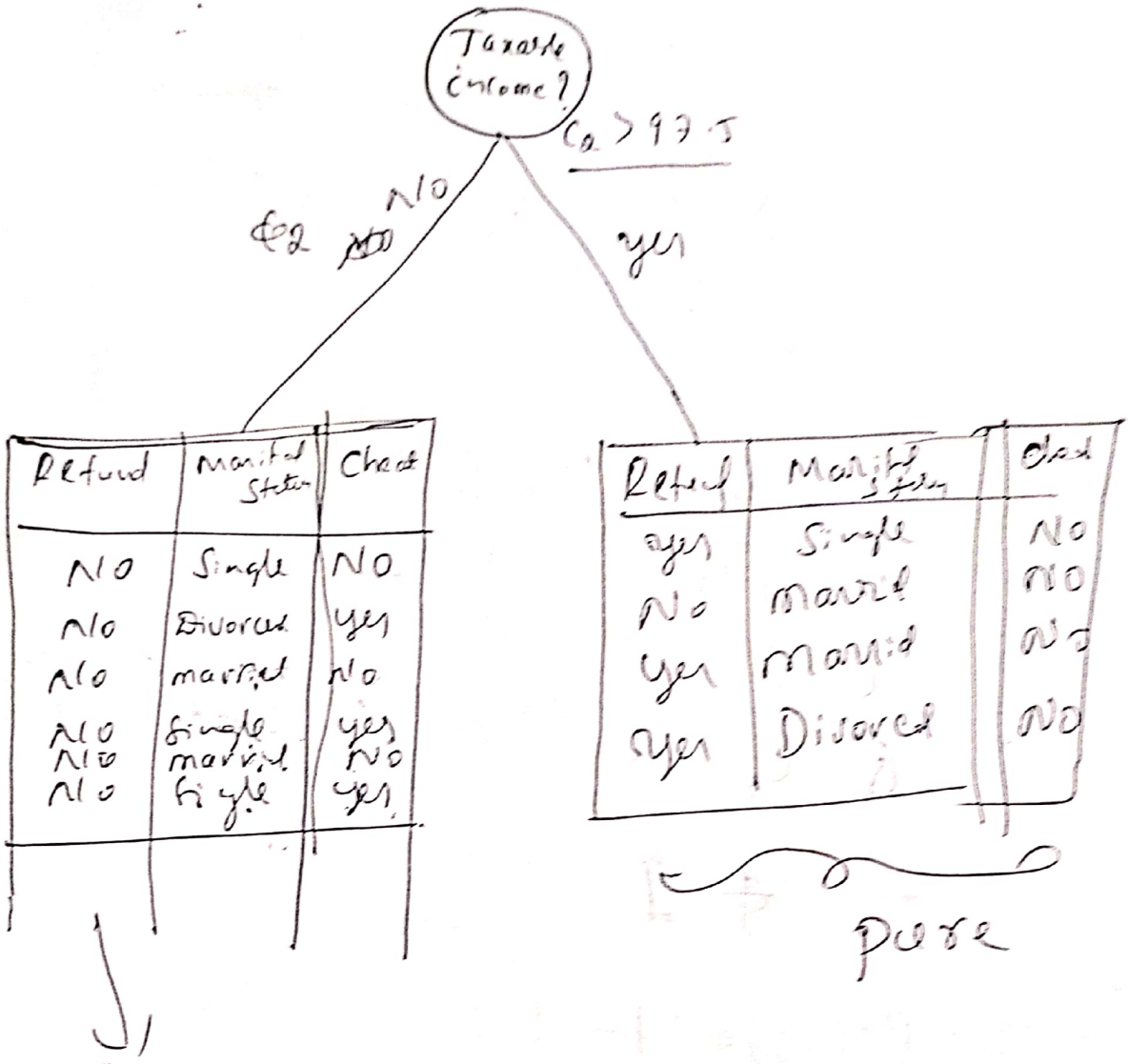
$$= \left(\frac{3}{10}\right)\left[-\left(\frac{0}{3}\right)\log_2\left(\frac{0}{3}\right) - \left(\frac{3}{3}\right)\log_2\left(\frac{3}{3}\right)\right] \quad (0)$$

$$+ \left(\frac{7}{10}\right)\left[-\left(\frac{3}{7}\right)\log_2\left(\frac{3}{7}\right) - \left(\frac{4}{7}\right)\log_2\left(\frac{4}{7}\right)\right]$$

$$= 0.68$$

$$Gain(Refund) = 0.88 - 0.68$$

$$= 0.20$$



Since it is impure, Split

$$\text{Info}(D) = -\sum_{i=1}^{2} p_i \log_2 p_i$$

$$= -\left[\frac{3}{6}\log_2\left(\frac{3}{6}\right) + \left(\frac{3}{6}\right)\log_2\left(\frac{3}{6}\right)\right]$$

$$= 1$$

$$Info(D)_{Refund} = \sum_{j=1}^{v} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{|D_{No}|}{|D|} Info(D_{No})$$

$$= \frac{6}{6}\left[-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right]$$

$$= 1$$

$$Gain(Refund) = 1 - 1$$

$$= 0$$

$$Info(D)_{marital\ Status} = \sum_{j=1}^{v} \frac{|D_j|}{|D|} Info(D_j)$$

$$= \frac{3}{6}\left[-\frac{2}{3}\log_2(2/3) - \left(\frac{1}{3}\right)\log_2\left(\frac{1}{3}\right)\right]$$

$$+ \frac{2}{6}\left[-\frac{0}{2}\log_2(0/2) - (2/2)\log_2(2/2)\right]$$

$$+ \frac{1}{6}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log\left(\frac{0}{1}\right)\right]$$

$$= 0.4591$$

Eg Gain (marital Status)

$= 1 - 0.4591$

$= 0.540.$