

AML ASSIGNMENT #02

BELAL AHMED KHAN ; K200183 ;

QUESTION NO 01

Date: _____

Creating a decision tree:-

i) Using CART & C4.5 Method:-

$$\text{info}([7, 6]), \text{entropy}(7/13, 6/13) = -\frac{7}{13} \log_2(7/13) - \frac{6}{13} \log_2(6/13)$$

$$\text{info}([7, 6]) = 0.995$$

Info gain (Refund):-

For yes

$$\text{info}([3, 4]) = -\frac{3}{7} \log_2(3/7) - \frac{4}{7} \log_2(4/7) = 0.985$$

For no

$$\text{info}([4, 2]) = -\frac{4}{6} \log_2(4/6) - \frac{2}{6} \log_2(2/6) = 0.918$$

$$\text{info}([3, 4], [4, 2]) = \frac{7}{13} (0.985) + \frac{6}{13} (0.918) = 0.954$$

$$\text{Info gain (Refund)} = 0.995 - 0.954$$

$$\text{Info gain (Refund)} = 0.041$$

Info gain (Status)

For single:-

$$\text{info}([2, 3]) = -\frac{2}{5} \log_2(2/5) - \frac{3}{5} \log_2(3/5) = 0.971$$

For married:-

$$\text{info}([2, 2]) = -\frac{2}{4} \log_2(2/4) - \frac{2}{4} \log_2(2/4) = 1$$

For divorced:-

$$\text{info}([3, 1]) = -\frac{3}{4} \log_2(3/4) - \frac{1}{4} \log_2(1/4) = 0.811$$

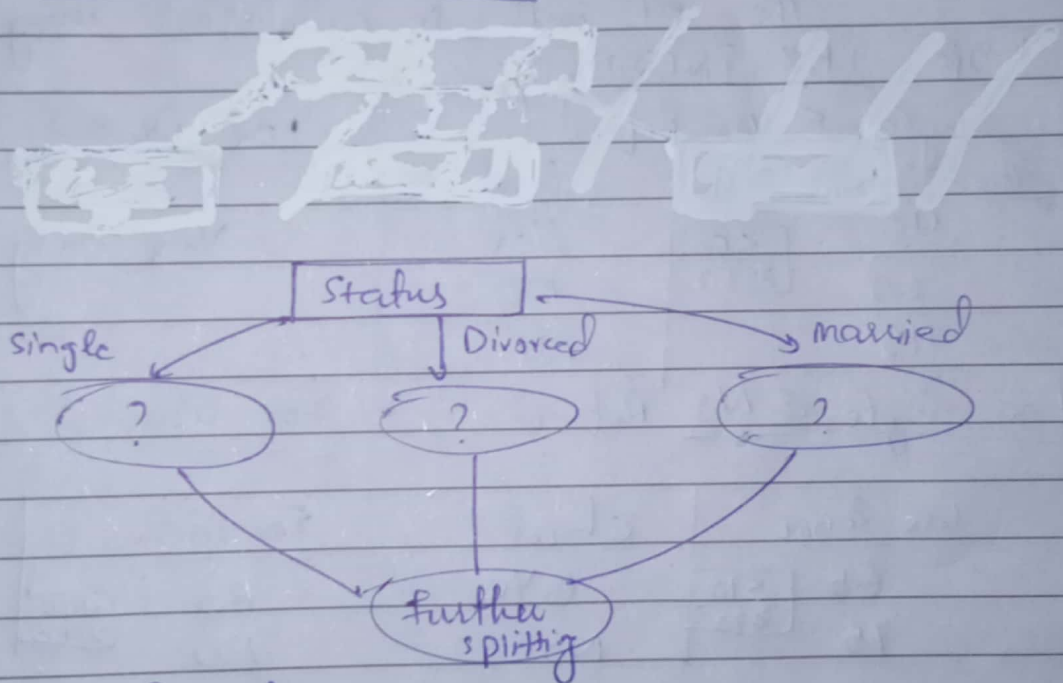
Date: _____

$$\text{Info}([2,5], [2,2], [3,1]) = \frac{5}{13} (0.921) + \frac{4}{13} (1) + \frac{4}{13} (0.811) = 0.931$$

$$\text{Infogain}(\text{Status}) = 0.995 - 0.931$$

$$\boxed{\text{Infogain}(\text{Status}) = 0.064}$$

We select (Status)



For Single branch:

For refund:

For yes:

$$\text{info}([1,2]) = 0.918$$

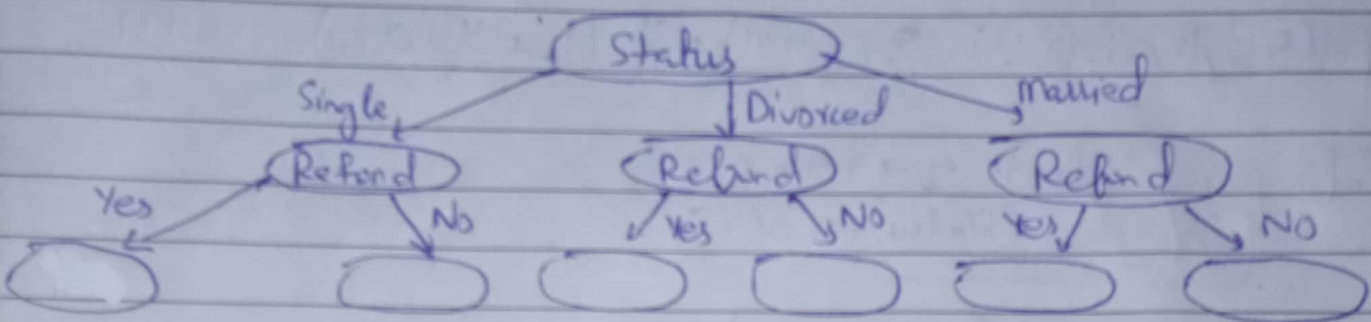
For no:

$$\text{info}([1,1]) = 1$$

$$\text{info}([1,2], [1,1]) = \frac{3}{5} (0.918) + \frac{2}{5} (1) = 0.9508$$

$$\boxed{\text{Infogain}[\text{refund}] = 0.0442}$$

Date: _____



The rest will be calculated using CART
FOR TAX INCOME:

For Single & Yes Refund

Tax Income	cheat
40	Yes
60	No
115	No

For Single & NO Refund

Tax Income	cheat
65	Yes
75	No

For Divorced & Yes Refund

Tax Income	cheat
80	No
130	Yes

For Divorced & NO Refund

Yes only

For Married & Yes Refund

Tax Income	cheat
80	No
130	Yes

For married & NO Refund

Tax Income	cheat
80	No
125	Yes

Date: _____

Final Tree will be

