NAME :- KRUNAL RANK

ROLL NO :- U18CO081

CLASS :- BTECH 4TH YEAR

SEMESTER :- 7

DIVISION :- B

Ans 1:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

$P = TP + FN$

$N = TN + FP$

$$RHS = \text{Sensitivity} \times \frac{P}{P+N} + \text{Specificity} \times \frac{N}{P+N}$$

$$= \frac{TP}{TP+FN} \times \frac{(TP+FN)}{(TP+TN+FP+FN)} + \frac{TN}{TN+FP} \times \frac{(TN+FP)}{(TP+TN+FP+FN)}$$

$$= \frac{TP+TN}{TP+TN+FP+FN}$$

= Accuracy

= LHS (Hence, proved).

Ans 1:

Magazine Promotion		Watch Promotion		Credit Insurance	Gender	
Yes	No	Yes	No	Yes	No	Male Female
6	4					

Life Insurance Promotion

Yes	4	1
No	2	3

Ratio (Yes)

Ratio (No)

Ans 2:

		Life Insurance - Promotion	
		Y	N
Magazine Promotion	Y	4	2
	N	2	3
P(MP | LIP)		4/5	2/5
		1/5	3/5
Watch Promotion	Y	4	1
	N	1	4
P(WP | LIP)		4/5	1/5
		1/5	4/5
Credit Card Insurance (CCI)	Y	3	1
	N	2	4
P(CCI | LIP)		3/5	1/5
		2/5	4/5
Gender	M	2	4
	F	3	1
P(Gender | LIP)		2/5	4/5
		3/5	1/5

a) X = {MP=Y, WP=Y, CCI=N, G=F}

$$P(LIP=Y|X) = \frac{P(X|LIP=Y) \cdot P(LIP=Y)}{P(X)}$$

$$P(X|LIP=Y) = P(MP=Y|LIP=Y) \times P(WP=Y|LIP=Y) \times P(CCI=N|LIP=Y) \times P(G=F|LIP=Y)$$

$$= 4/5 \times 4/5 \times 2/5 \times 3/5$$

$$= 96/625$$

$$P(LIP=Y|X) = \frac{96/625 \times 1/2}{P(X)} = \frac{48/625}{P(X)}$$

$$P(LIP=N|X) = \frac{P(X|LIP=N) \cdot P(LIP=N)}{P(X)}$$

$$P(X|LIP=N) = P(MP=Y|LIP=N) \times P(WP=Y|LIP=N) \times P(CCI=N|LIP=N) \times P(Gender=F|LIP=N)$$

$$= 1/5 \times 2/5 \times 1/5 \times 4/5$$

$$= \frac{8}{625}$$

$$P(LIP=N|X) = \frac{8/625 \times 1/2}{P(X)} = \frac{4/625}{P(X)}$$

Since $P(LIP=Y|X) > P(LIP=N|X)$

LIP=Y

b) If gender is unknown, LIP=Y

Ans 3:

$$H(S) = -\left(\frac{4}{10}\log_2\left(\frac{4}{10}\right) + \frac{3}{10}\log_2\left(\frac{3}{10}\right) + \frac{3}{10}\log_2\left(\frac{3}{10}\right)\right)$$

$$= 1.5709$$

For attribute Gender,

	Male	Female
Bus	III	I
Car	I	II
Train	I	II

$$H(\text{Gender} = \text{Male}) = -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{1}{5}\log_2\left(\frac{1}{5}\right) + \frac{1}{5}\log_2\left(\frac{1}{5}\right)\right)$$

$$= 1.3709$$

$$H(\text{Gender} = \text{Female}) = -\left(\frac{1}{5}\log_2\left(\frac{1}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$$

$$= 1.5219$$

$$H(\text{Gender}) = \frac{1}{2} * 1.3709 + \frac{1}{2} \times 1.5219$$

$$= 1.4464$$

$$IG = 1.5709 - 1.4464 = 0.1245$$

For attribute Car Ownership,

	0	1	2
Bus	II	II	
Car		I	II
Train	I	II	

$$H(\text{Car Ownership} = 0) = -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$$

$$= 0.9234$$

$$H(\text{Car Ownership} = 1) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{1}{5}\log_2\left(\frac{1}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$$

$$= 1.5219$$

$H(\text{Car Ownership}=2) = -(1\log_2(1)) = 0$

$H(\text{Car Ownership}) = \frac{3}{10} \times 0.9234 + \frac{5}{10} \times 1.5219 + 0$

$= 1.0379$

$IG = 1.5709 - 1.0379 = 0.5330$

For Attribute Travel Cost,

	Cheap	Standard	Expensive
Bus	1111		
Car			111
Train	1	11	

$H(\text{Travel Cost}=\text{Cheap}) = -\left(\frac{4}{5}\log_2\left(\frac{4}{5}\right) + \frac{1}{5}\log_2\left(\frac{1}{5}\right)\right)$

$= 0.7219$

$H(\text{Travel Cost}=\text{Standard}) = 0$

$H(\text{Travel Cost}=\text{Expensive}) = 0$

$H(\text{Travel Cost}) = 0.3609$

$IG = 1.5709 - 0.3609 = 1.21$

For attribute Income level,

	Low	Medium	High
Bus	11	11	
Car		1	11
Train		111	

$H(\text{Income level}=\text{Low}) = -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) = 1 = 0$

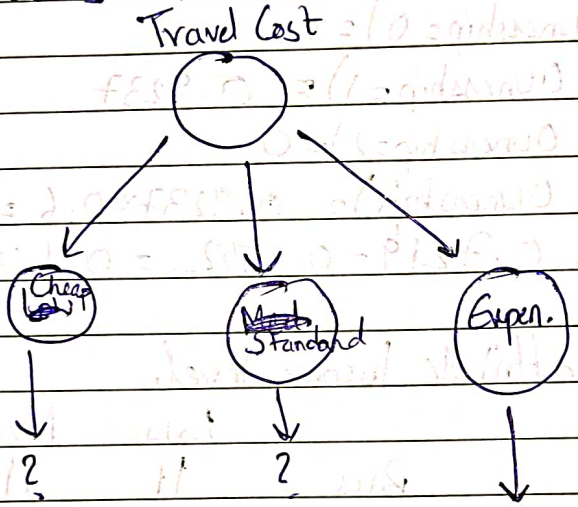
$H(\text{Income level}=\text{Medium}) = -\left(\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{6}\log_2\left(\frac{1}{6}\right)\right) = 1.4586$

$H(\text{Income level}=\text{High}) = 0$

H(Income level) = 1.4586 × 0.6
= 0.8751

IG = 1.5709 − 0.8751 = 0.6958

Hence, Highest IG is for Attribute Travel Cost.
So root node is Travel Cost.



Lets solve for Travel Cost = Cheap

H(Travel Cost = Cheap) = 0.7219

Now, for attribute Gender,

	Male	Female
Bus	111	1
Car		
Train		1

H(Gender = Male) = 0

H(Gender = Female) = 1

H(Gender) = 0.4

IG = 0.3219

For attribute Car Ownership,

	0	1	2
Bus	11	11	
Car			
Train		1	

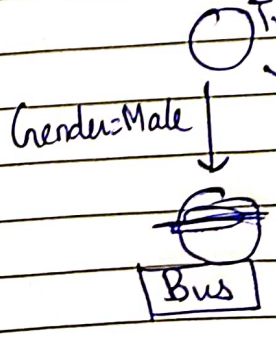
H(Car Ownership=0) = 0
H(car Ownesship=1) = 0.9237
H(car Ownership=2) = 0
H(Car Ownership) = 0.9237 × 0.6 = 0.5542
IG = 0.7219 - 0.5542 = 0.1676

For ~~Incor~~ attribute Income level,

	Low	Medium	High
Bus	11	11	
Car			
Train			1

H(Income level = low) = 0
H(Income level = medium) = 0.9237
H(~~Car~~ Income level = high) = 0
H(Income level) = 0.5542
IG = 0.7219 - 0.5542 = 0.1676

Hence, Highest IG is for Gender attribute.
So,



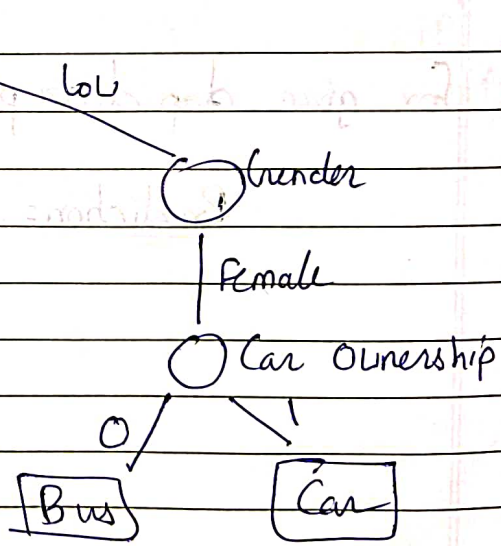
Now, H(Travel Cost = Low | Gender = Female) = 1

For attribute Car Ownership

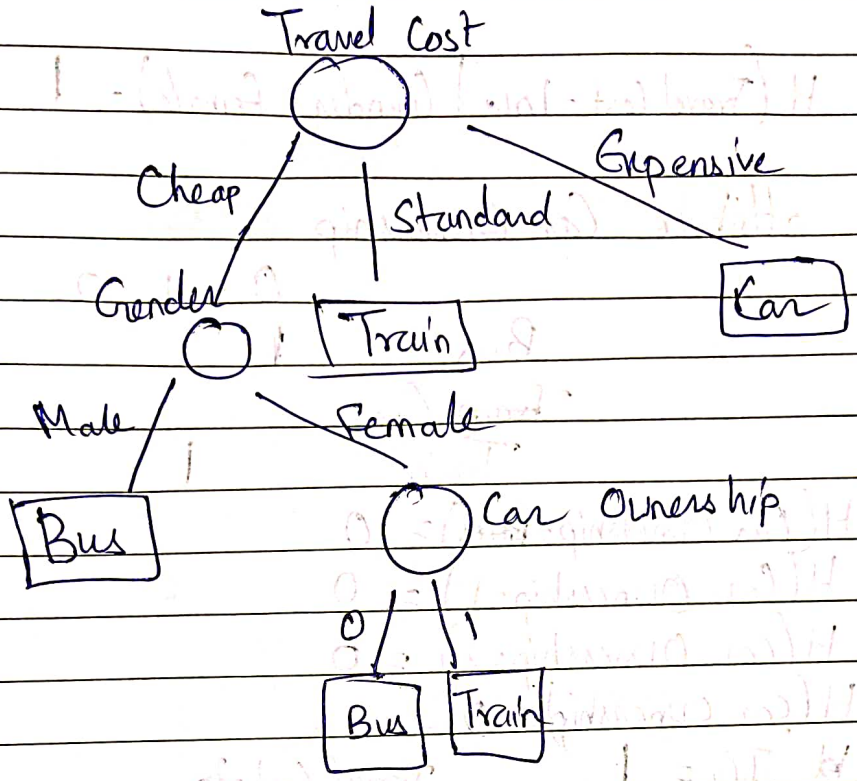
	0	1	2
Bus	1		
~~Train~~ Car			
Train		1	

H(Car Ownership = ~~Bus~~ 0) = 0
H(Car Ownership = 1) = 0
H(Car Ownership = 2) = 0
H(Car Ownership) = 0
IG = 1

Hence



Now, let us calculate sub tree for Travel Cost = Standard
For attribute Travel Cost = Standard, all outputs are Train. Hence, the constructed Decision Tree is as follows:-



for given ~~dog~~ data point <Female, 2, Standard, High>

Prediction: Train.