Decision tree assignment.

1) chest pain decision tree for heart attack.

id	chestpain	male	Smokers	Exercise	Heart Attack.
1	yes	yes	no	yes	yes
2	yes	yes	yes	no	yes
3	NO	no	yes	no	yes
4	NO	yes	no	yes	no
5	yes	no	yes	yes	yes
6	NO	yes	yes	yes.	no.

step 1:- Find entropy of training set.

$$\text{Entropy} = E[(4,-2)] = (-4/6)\log_2(4/6) + (-2/6)\log_2(2/6).$$

$$= 0.918295.$$

step 2 :- Information gain for chest pain.

$$G(S, \text{chest pain}) = 0.918295 - [(3/6)(-\log_2(1)) + (3/6)((-1/3)\log_2(1/3) + (-2/3)\log_2(2/3))]$$

$$= 0.4591.$$

step 3:- Information gain for male attribute.

$$G(S, \text{male}) = 0.918295 - ((4/6)((-2/4)\log_2(2/4) + (-2/4)\log_2(2/4)) + (2/6)(-\log_2(1)))$$

$$= 0.251.$$

step 4:- Information gain for smokers attribute.

$G(S, \text{smokers}) = 0.918295 - ((4/6)((-3/4)\log_2(3/4) + (-1/4)\log_2(1/4)) + (2/6)((-1/2)\log_2(1/2) + (-1/2)\log_2(1/2)))$

$= 0.0441.$

step 5 :- Information gain for exercise attribute.

$G(S, \text{exercise}) = 0.918295 - ((4/6)((-2/4)\log_2(2/4) + (-2/4)\log_2(2/4)) + (2/6)((-1)\log_2(1))$

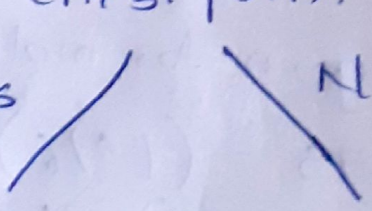
$= 0.2516.$

step 6 :-

	Information gain.
Chest pain.	0.4591
mole	0.251629
smokes	0.0441
Exercise	0.25169

chest pain will be root.

step 7:- Construct tree



Divide data set based on attribute.

		Heart Attack.	
		Yes	No
chest pain	Yes	3	0
	No	1	2

chestpain	Male	smokes	Excersise	Heart Attack.
① Yes	Yes	No	Yes	Yes
Yes	Yes	Yes	No	Yes
Yes	No	Yes	Yes	Yes

chestpain	male	Smokes	Excersise	Heart Attack
② No	No	Yes	No	Yes
No	Yes	No	Yes	No
No	Yes	Yes	Yes.	No.

Eubopy for ①.

$E(s) = E((3)) = (-1)\log_2(1) = 0$

Eubophy for ②

$E(s) = E[(1,-2)] = \left(-\frac{1}{3}\log_2(\frac{1}{3}) + (-\frac{2}{3})\log(\frac{2}{3})\right)$

$= 0.91829.5.$

Step 8:- Information gain for all attributes wrt data set ②.

→ For male

$G(S, male) = 0.9182 - ((2/3)([-2/2)\log_2(2/2)] + (1/3)((-1)\log_2(1))$

$= 0.9182$

For smokes

$G(S, \text{smokes}) = 0.9182 - ((2/3)((-1/2)\log_2(1/2), (-1/2)\log_2(1/2)) + (1/3)(0))$

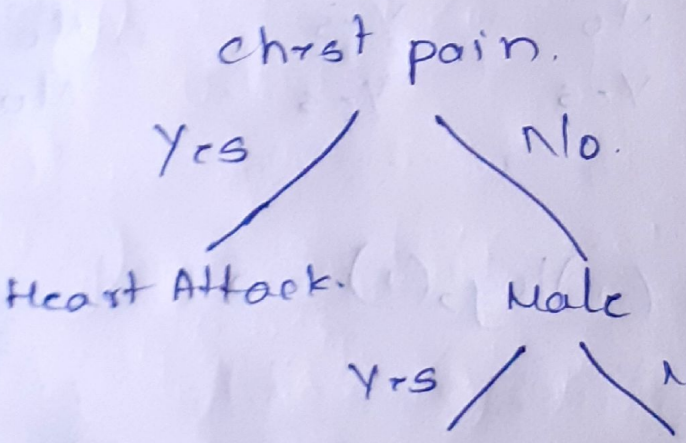
$= 0.2515.$

For Exercise

$G(S, \text{Exercise}) = 0.9812 - ((2/3)(0) + (1/3)(0))$

$= 0.9182.$

	Info gain.
Male	0.9182
Smokes	0.2515
Exercise.	0.9182.



Step 2:- Divide data set.

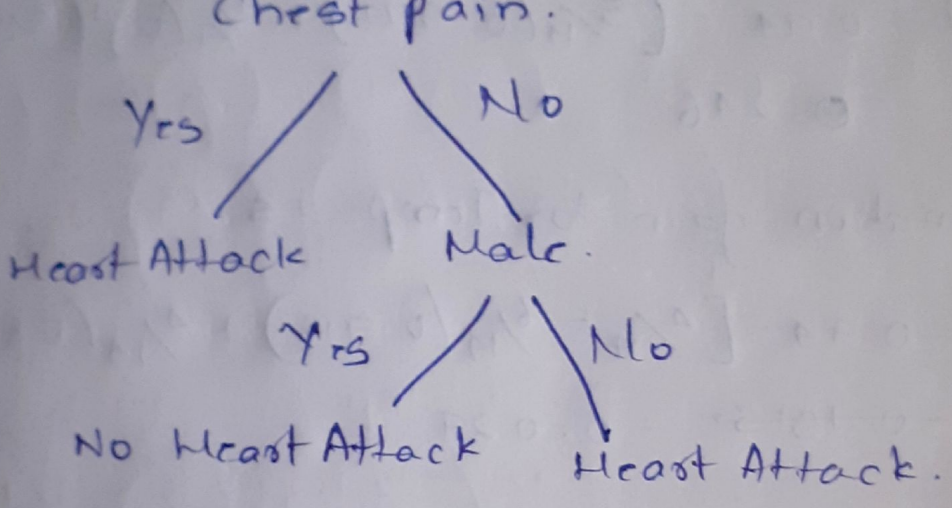
	Male	smokes	Exercise	Heart Attack.
①	Yes	No	Yes	No
	Yes	Yes	Yes	No

	Male	smokes	Exercise	Heart Attack
②	No	Yes	No	Yes

Entropy for ① is 0.

Entropy for ② is 0.

Final tree.



② Play Tennis Decision Tree.

Day	Outdoor	Temp	Humidity	Wind	Play Tennis.
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Norm	Weak	Yes
6	Rain	Cool	Norm	Strong	No
7	Overcast	Cool	Norm	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Norm	Weak	Yes
10	Rain	Mild	Norm	Weak	Yes
11	Sunny	Mild	Norm	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Norm	Weak	No.
14	Rain.	Mild	High	Strong	

1) step 1:- Find Entropy of data set

$E(S) = E([9, 5-]) = (-9/14)\log_2(9/14) + (-5/14)\log_2(5/14)$

$= 0.94$

step 2:- Information gain for outlook is

$G(S, outlook) = 0.94 - [5/14(0.971) + 4/14(0) + (5/14)](0.97$

$= 0.246$

step 3:- Information gain for Temp is.

$G([S, Temp) = 0.94 - [4/14 + 6/14(0.918) + 4/14(0.811)]$

~~0.1515~~ $= 0.029$.

step 4:- Information gain for Humidity is

$G(S, Humidity) = 0.94 - (7/14 \times 0.985 + 7/14(0.592))$

$= 0.1515$.

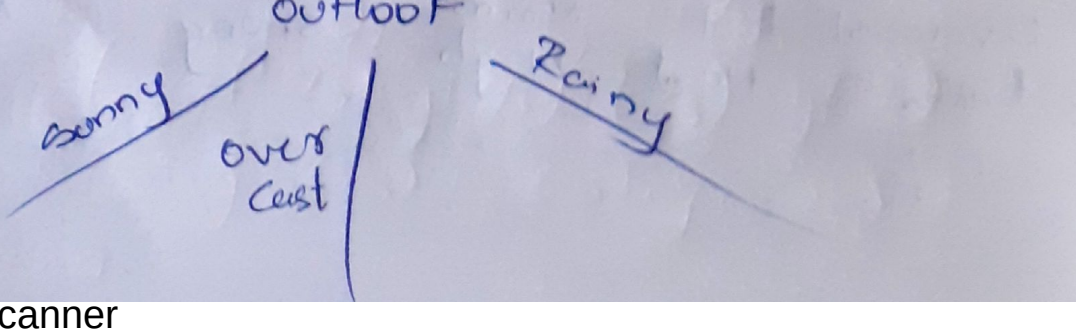
step 5:- Information gain for windy is

$G(S, windy) = 0.94 - [8/14 * 0.811 + 6/14 * 1]$

$= 0.048$.

	Information gain
Outlook	0.247
Temp	0.029
Humidity	0.152
Windy	0.048

outlook will be root

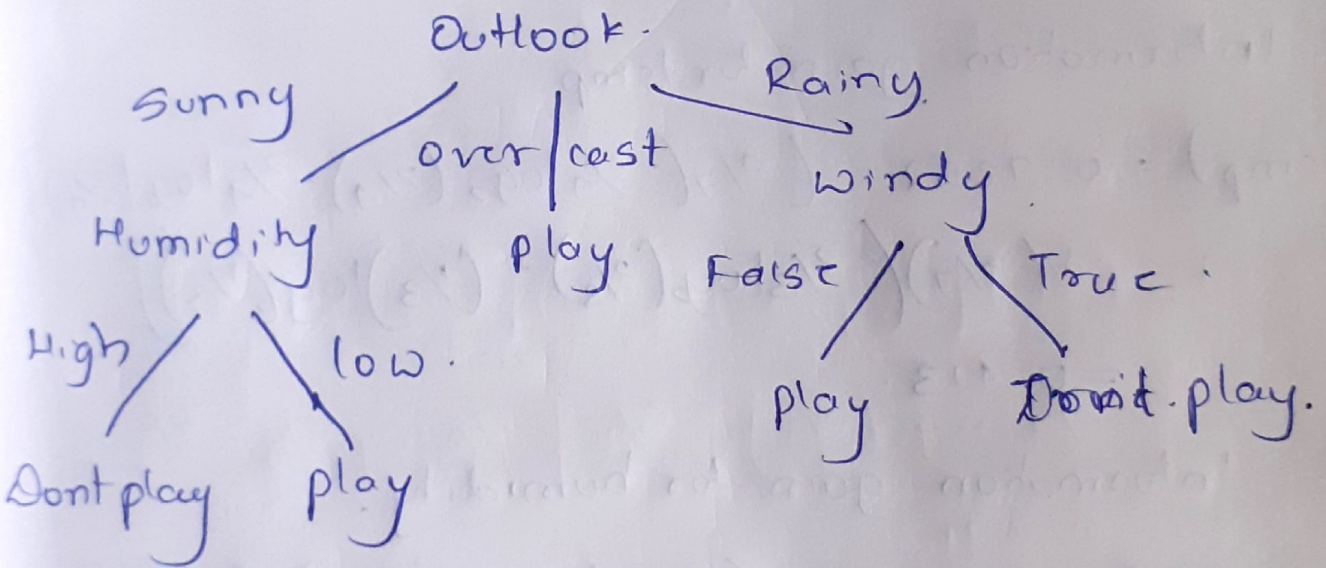


Entropy and Information gain after dividing the set.

	Information gain.
Temperature	0.571
Humidity	0.971
Windy	0.020.

Also entropy of overcast set is 0.

∴ Overcast will be leaf.



b) For day 1 to 7.

Day	Outlook	Temp	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	rainy	Mild	High	False	Yes
5	rainy	Cool	Normal	false	Yes
6	rainy	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes

Step 1:- Entropy of Data set.

$E(S) = E((4,3^-)) = (-4/7 \log_2 4/7) + (-3/7)\log_2(3/7)$

$= 0.9852.$

Step 2:- Information gain for outlook.

$G(S, \text{outlook}) = 0.9852 - (2/7(0) + (2/7)(0) + (3/7)$

$(-2/3 \log_2(2/3) + (-1/3)\log_2(1/3))$

$= 0.591$

Step 4:- Information gain for temp

$G[S, \text{temp}) = 0.9852 - (4/7(-2/4 \log_2(2/4) - 2/4 \log_2(2/4)$

$+ (3/7)(-2/3 \log_2(2/3) - (1/3)\log_2(1/3))]$

$= 0.413.$

Step 5:- Information gain for humidity.

$G(S, \text{Humidity}) = 0.9852 - (4/7(-1/2 \log(1/2) - (1/2)\log(1/2)$

$+ 3/7(-2/3 \log(2/3) - 1/3 \log(1/3))$

$= 0.9852 - (4/7(-\log(1/2)) + 3/7(-2/3 \log(2/3)$

$- 1/3 \log(1/3))$

$= 0.0202.$

Step 6:- Information gain for windy

$= 0.9852 - (3/7(-2/3 \log(2/3) - (1/3)\log(1/3)) + 4/7$

$(-0.75 \log(0.75) - 0.25 \log(0.25))$

$= 0.089.$

	Information gain
Outlook	0.591
temp	0.413
Windy	0.089
Humidity	0.020

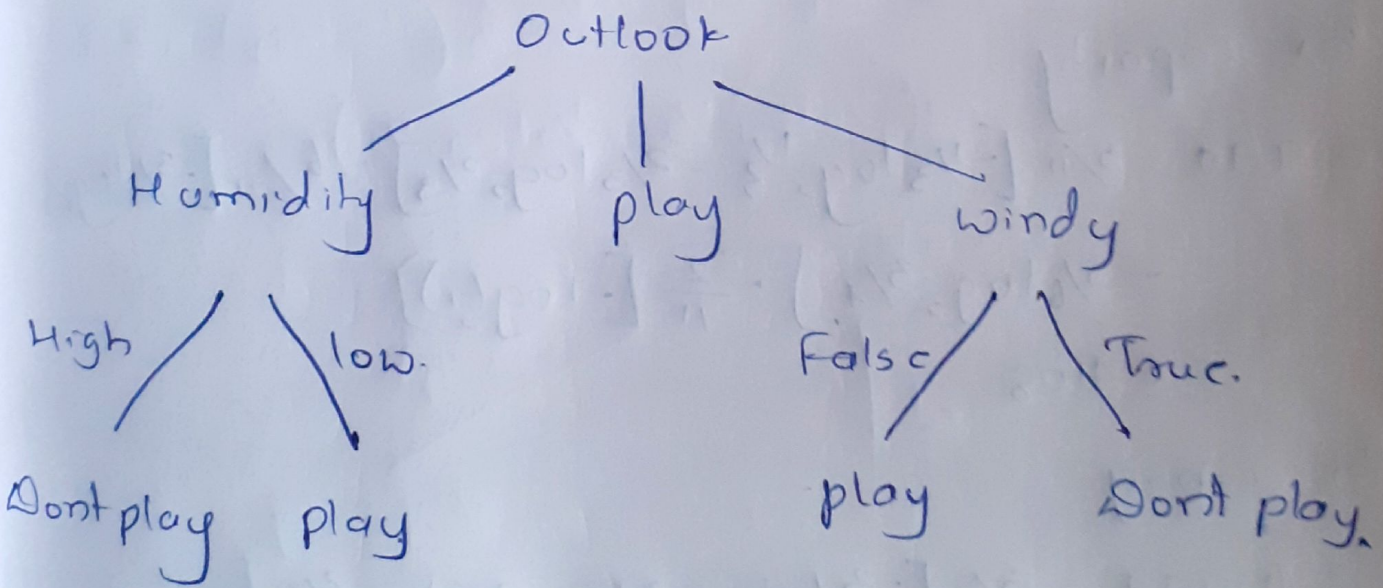
∴ outlook will be root

Step 1:- Find information gain based on outlook.

	Information gain.
Temp	0.98
Humidity.	0.71
Windy	0.017.

Entropy of overcast is 0.

∴ it will be leaf



a) Entropy $= E([6,-5]) = -(6/11)\log_2(6/11) + (5/11)\log_2(5/11) = 0.994$.

IG for price $= [0.994 - (4/11)[-2/4\log_2(2/4) - 2/4\log_2(2/4)] - (4/11)[-2/4\log_2(2/4) - 2/4\log_2(2/4)] - 3/11[-2/3\log_2 2/3 - 1/3\log_2 1/3]$

$= 0.0167$.

IG for maintanance.

$= 0.994 - (2/11)[-\log_2(1)] - (4/11)[-2/4\log_2(2/4) - 2/4\log_2(2/4)] - 5/11[-2/5\log_2 2/5 - 3/5\log_2 3/5]$

$= 0.591$

IG for capacity

$= 0.994 - 3/11[-1/3\log_2 1/3 - 2/3\log_2 2/3] - 6/11[-3/6\log_2 3/6 - 3/6\log_2 3/6] - \frac{2}{11}[-\log(1)]$

$= 0.4108$

IG for airbag

$= 0.994 - 5/11[-3/5\log_2 3/5 - 2/5\log_2 2/5] - 6/11[-3/6\log_2 3/6 - 3/6\log_2 3/6]$

$= 0.0075$

∴ Maintainence is root

Maintonence

Low | Med | High

Profitable.

	Maintance	Price	Capacity	Airbag	Profitable.
(i)	Low	low	2	No	Yes
	Low	low	4	No	Yes.
(ii)	Med	low	4	Yes	No
	Med	med	4	No	No
	Med	med	4	Yes	Yes
	Med	high	4	Yes	Yes.
(iii)	High	low	4	No	No
	High	med	2	yes	No
	High	med	5	No	yes
	High	high	2	yes	No
	High	high	5	Yes	yes

(i) Entropy = $(E[3]) = -1 \log_2 1 = 0$ Leaf.

(ii) Entropy = $E([2,-2]) = -2/4 \log_2 2/4 - 2/4 \log 2/4 = 1$.

→

IG for price = $1 - 4/4(-\log_2) - 2/4(-1/2\log_2(1/2) - 1/2\log_2 1/2) - 1/4(-\log_2(1))$

$= 0.5$

IG for capacity $= 1 - 4/4(-2/4\log_2 2/4 - 2/4\log_2 2/4)$

$= 0$

IG for airbag $= 1 - 3/4[-2/3\log 2/3 - 1/3\log_2 1/3] - 1/4[-\log(1)]$

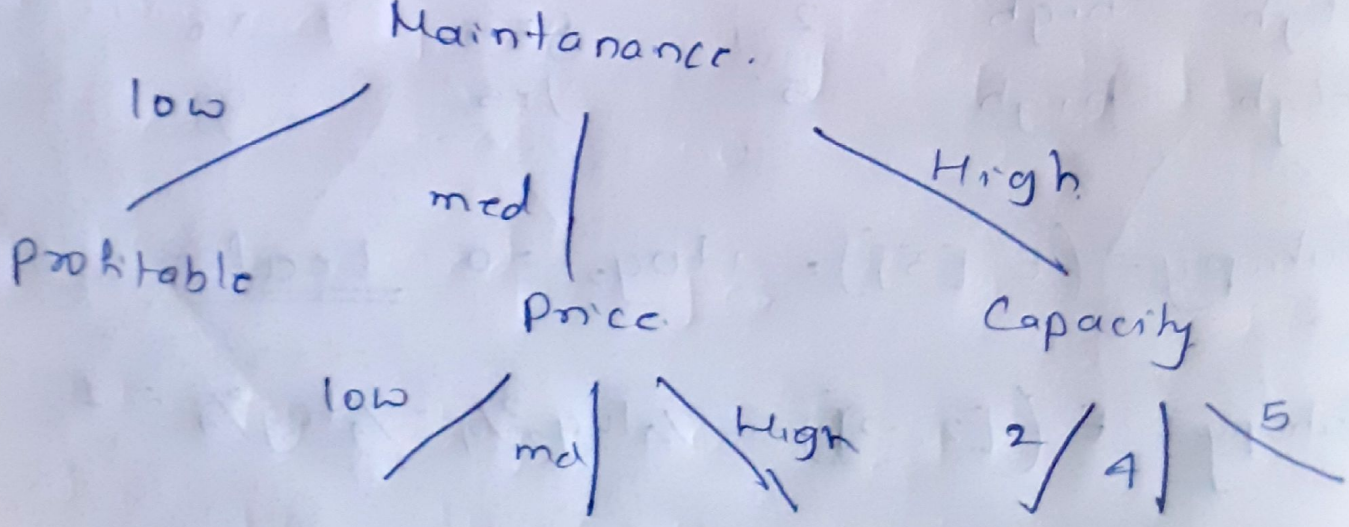
$= 0.3113$.

Prize ne price next Node.

Entropy $= E([2,3]) = \frac{-2}{5}\log_2 2/5 - 3/5\log_2 3/5$

$= 0.971$

IG for price $= 0.971 - 1/5(-\log_2(1)) - 2/5(-1/2\log_2 1/2 - 1/2\log_2 1/2) - 2/5(-1/2\log_2 1/2 - 1/2\log_2 1/2)$



a)

	Maintanance	Price	Capacity	Airbag	Profitable
(i)	med	low	4	yes	No
(ii)	med	med	4	No	No
	med	med.	4.	Yes.	Yes.
(iii)	med	High	4	yes	yes.

b)

Maintanance	Capacity	Price	airbag	profitable.
High	2	med	yes	no
High	2	High	yes	no
High	4	low	no	no
High	5	med	yes	yes
High	5.	High	yes	yes.

a) (i) Entropy $= E[-1] = -1 \log_2 1 = 0$ Leaf.

(ii) Entropy $= E[1,-1] = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1$.

IG for Capacity $= 1 - 1(-1/2 \log_2 1/2 - 1/2 \log_2 1/2)$

$= 0$.

IG for airbag $= 1 - 1/2(-\log_2(1)) - 1/2(-\log_2(1))$

$= 1$.

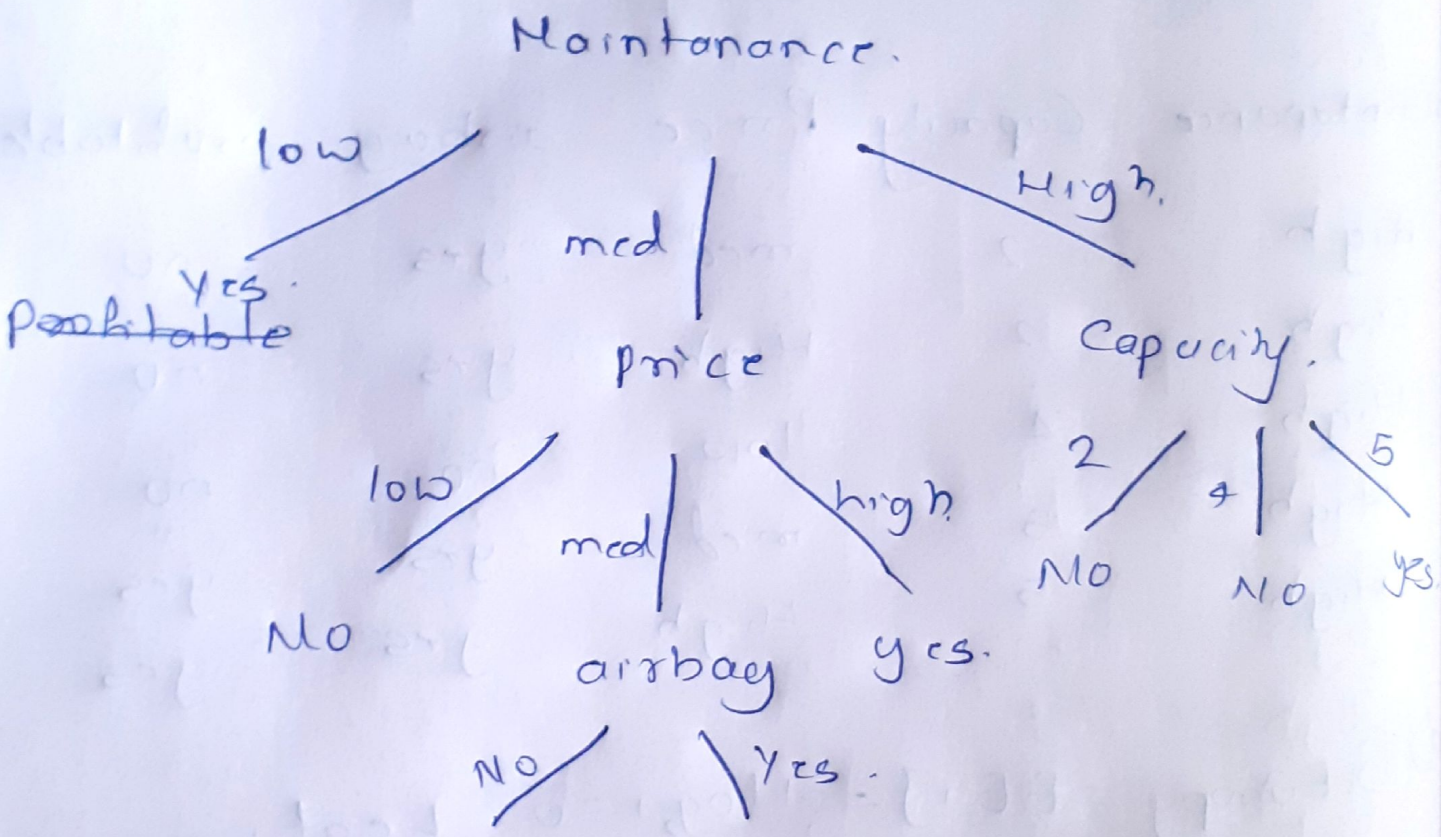
Airbag next child.

(iii) Entropy = $E[1] = -1\log_2(1) = 0$ Leaf.

b.) (i) Entropy = $-1\log_2(1) = 0$. Leaf

(ii) Entropy = $-1\log_2(1) = 0$ Leaf

(iii) Entropy = $-1\log_2(1) = 0$ Leaf.



	Maintenance	price	airbag	Capacity	Probitable
(i)	med	med	no	4	no
(ii)	med	med	yes	4	yes

(i) Entropy = $-\log_2 1 = 0$. Leaf

(ii) Entropy = $-\log_2 1 = 0$ Leaf.