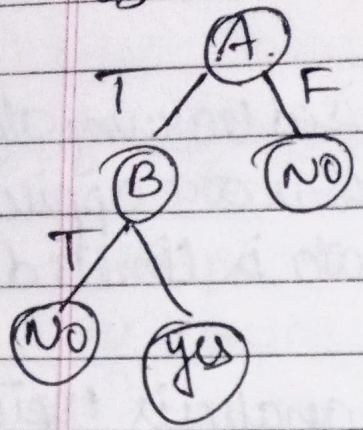
Assignment - 2 Decision tree.

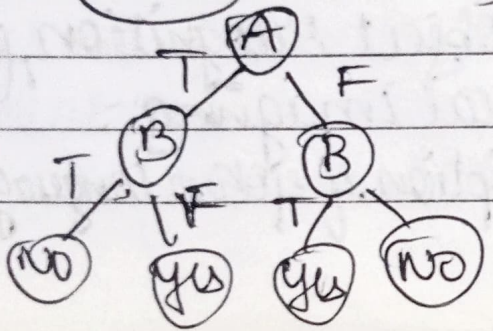
1) Develop the decision tree for the following

a) $A \wedge \neg B$



A	B	¬B	(A ∧ ¬B)
T	T	F	F
T	F	T	T
F	T	F	F
F	F	T	F

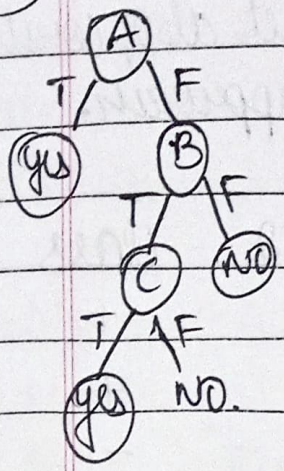
2) A XOR B $(A \oplus B)$



A	B	A XOR B
T	T	F
T	F	T
F	T	T
F	F	F

PAGE NO.
DATE / /

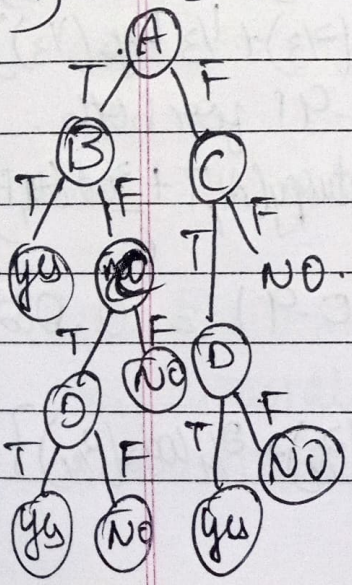
2) $A \vee [B \wedge C]$

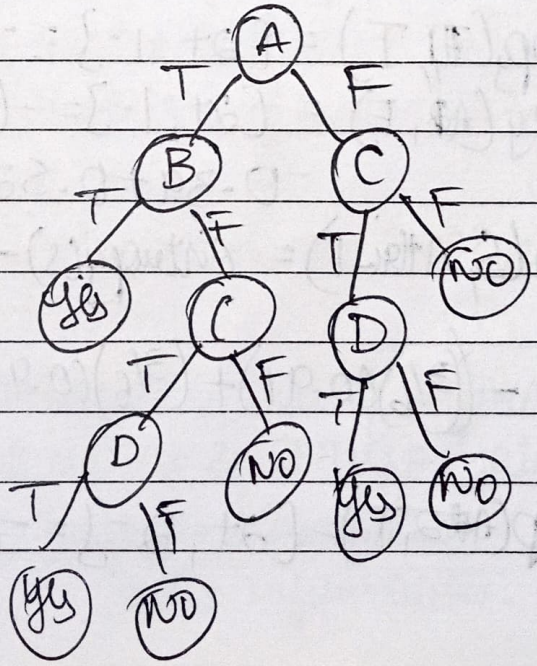


A	B	C	$B \wedge C$	$A \vee [B \wedge C]$
F	F	F	F	F
F	F	T	F	F
F	T	F	F	F
F	T	T	T	T
T	F	F	F	T
T	F	T	F	T
T	T	F	F	T
T	T	T	T	T

A	B	C	D	$A \wedge B$	$C \wedge D$	$[A \wedge B] \vee [C \wedge D]$
F	F	F	F	F	F	F
F	F	F	T	F	F	F
F	F	T	F	F	F	F
F	F	T	T	F	T	T
F	T	F	F	F	F	F
F	T	F	T	F	F	F
F	T	T	F	F	F	F
F	T	T	T	F	T	T
T	F	F	F	F	F	F
T	F	F	T	F	F	F
T	F	T	F	F	F	F
T	F	T	T	F	T	T
T	T	F	F	T	F	T
T	T	F	T	T	F	T
T	T	T	F	T	F	T
T	T	T	T	T	T	T

4) $[A \wedge B] \vee [C \wedge D]$





2) Develop the decision tree using ID3 algorithm and Gini index methods for the following data set. Also, construct the decision tree using gain ratio approach.

Instance	Attribute-1	Attribute-2	Class
1	T	T	+
2	T	T	+
3	T	F	−
4	F	F	+
5	F	T	−
6	F	T	−

1) ID3 Algorithm (using Entropy and Information Gain)

Entropy $(S) = \sum_{v=All} \frac{|S_v|}{|S|}$ i.e $[3+, 3-]$

$\Rightarrow -(P_\oplus \log_2 P_\oplus + P_\ominus \log_2 P_\ominus) \Rightarrow -\left(\frac{3}{6}\log_2\left(\frac{3}{6}\right) + \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right)$

" 1 " $\equiv -(0.5(-1) + 0.5(-1))$

Entropy $(Att_1, T) = [2+, 1-] = -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$

Entropy $(Att_1, F) = [2+, 1-] = -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$

$= 0.39 + 0.52 = 0.91$ for both.

Gain $(S, Attr_1) = \text{Entropy}(S) - \left(\frac{3}{6}\text{entropy}(Att_1, T) + \frac{3}{6}\text{entropy}(Att_1, F)\right)$

$1 - \left(\left(\frac{3}{6}\right)(0.91) + \left(\frac{3}{6}\right)(0.91)\right) = 0.91 \approx 0.082$

Entropy $(Att_2, T) = [2+, 2-] = -\left(\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right)$

$= 1$

Entropy $(Att_2, F) = [1+, 1-] = -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$

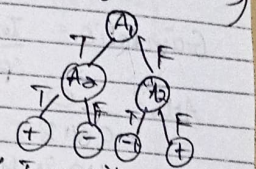
$= 1$

Gain $(S, Attr_2) = \text{entropy}(S) - \left(\frac{4}{6}\text{entropy}(Att_2, T) + \frac{2}{6}\text{entropy}(Att_2, F)\right)$

$1 - (0.66 + 0.33) = 0$

Gain $(Att_1) = 0.082$ Gain $(Att_2) = 0$

Att 1 will be the root node.



using Gini Index is a measure of impurity used to decide splits in decision tree.

Gini Index $\Rightarrow$ Gini $(S) = 1 - \sum P_i^2$

Gini $(S) = 1 - \left(\left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right) = 1 - (0.25 + 0.25) = 0.5$

Gini $(S, T) = 1 - \left(\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right) = 1 - (0.44 + 0.11) = 0.44$.

Gini $(S, F) = 1 - \left(\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2\right) = 0.44$.

Gini $(Attr_1) = \frac{3}{6} * 0.44 + \frac{3}{6} \times 0.44 = 0.44$.

Gini $(Att_2, T) = 1 - \left(\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right) = 0.5$

Gini $(Att_2, F) = 1 - \left(\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right) = 0.5$

Gini $(Att_2) = \frac{4}{6} \times 0.5 + \frac{2}{6} 0.5 \quad 0.33 + 0.16 = 0.49 \approx 0.5$

Gini (Att_1) is less so it will be the root node.

Gain Ratio: method address the bias of information gain by normalizing it with a split information value

Gain Ratio $=$ $\frac{\text{Information Gain}}{\text{Split Information}}$.

Split Information (att1) $= -\left(\frac{3}{6}\log_2\left(\frac{3}{6}\right) + \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right) = 1$

Split Information (att2) $= -\left(\frac{4}{6}\log_2\left(\frac{4}{6}\right) + \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right) = 0.9$

Gain Ratio $=$ $\frac{\text{Info Gain}}{\text{Split Info}}$ $\quad$ Att1 $= \frac{0.082}{1} = 0.082$

Att2 $= \frac{0}{0.918} = 0$

Gain ratio is high for Attribute 1 so it will be the gain ratio.

Assignment-4

Develop a decision tree using Gini index.

Day	outlook	Temperature	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	Normal	weak	yes
6	Rain	cool	Normal	Strong	NO
7	Overcast	cool	Normal	Strong	yes
8	Sunny	Mild	High	weak	NO
9	sunny	cool	Normal	weak	yes
10	Rain	Mild	Normal	weak	yes
11	Sunny	Mild	Normal	Strong	yes
12	Overcast	Mild	High	Strong	yes
13	Overcast	Hot	Normal	weak	yes.
14	Rain.	Mild	High	Strong	NO