CS6375- Assignment-2

Decision Tree Induction

Part-1

2. Entropy of whole dataset $= \frac{5}{10} = E_1$

Instance	2, 3, 5, 7, 9	1, 4, 6, 8, 10
class	1, 0, 0, 0, 0	1, 0, 1, 1, 1

$$E = -p \log_2 p - (1-p) \log_2 (1-p)$$

$$E_2 = -\frac{1}{5} \log_2 \left(\frac{1}{5}\right) - \left(\frac{4}{5}\right) \log_2 \left(\frac{4}{5}\right)$$

$(X_1 = F)$

$$= 0.4643 + 0.2575$$

$$E_2 = 0.7218$$

$$E_3 = -\frac{4}{5} \log_2 \left(\frac{4}{5}\right) - \left(\frac{1}{5}\right) \log_2 \left(\frac{1}{5}\right)$$

$(X_1 = T)$

$$= 0.2575 + 0.4643$$

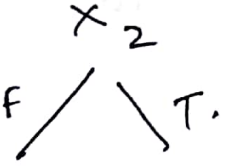
$$= 0.7218$$

$$IG_1 = E_1 - \frac{5}{10} E_2 - \frac{5}{10} E_3$$

$$= \frac{1}{2} - \frac{1}{2}(0.7218) - \frac{1}{2}(0.7218)$$

$= 0.5 - 0.3609 - 0.3609$

$IG_1 = \underline{-0.2218}$



Instances	1, 3, 4, 5, 8, 9, 10	2, 6, 7
Class	1, 0, 0, 0, 1, 0, 1	1, 1, 1

$E_1 = 1$

$E_{(X_2=F)} = -\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \left(\frac{4}{7}\right)\log_2\left(\frac{4}{7}\right)$

$= 0.5238 + 0.4613$

$= 0.9851$

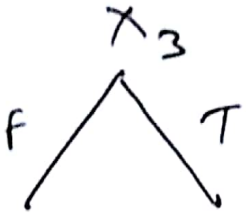
$E_{3(X_2=T)} = -1\log_2 1$

$= 0$

$IG_2 = E_1 - \frac{7}{10}E_2 - \frac{3}{10}E_3$

$= 1 - \frac{7}{10}(0.9851) - 0$

$IG_2 = = \underline{0.31043}$



Instances	1, 2, 3, 5, 6, 8, 9, 10	4, 7
class	1, 1, 0, 0, 1, 1, 0, 1	0, 0

$E_1 = 1$

$$E_{2\,(X_3=F)} = -\frac{5}{8}\log_2\left(\frac{5}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right)$$

$$= 0.42379 + 0.53063$$

$$= 0.95442$$

$$E_{3\,(X_3=T)} = 0 - 1\log_2 1$$

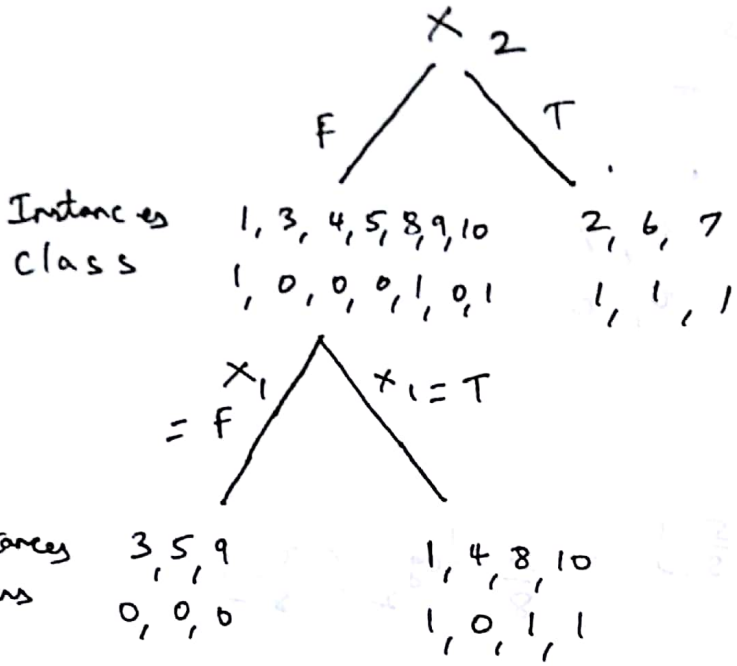
$$= 0$$

$$IG_3 = E_1 - \frac{8}{10}E_2 - \frac{2}{10}E_3$$

$$= 1 - 0.763536 - 0$$

$$= \underline{0.236464}$$

$\therefore$ X_2 has more I.G., we split based on X_2.



wkt, $E_{(X_2=F)} = 0.9851$

$E_{(X_2=F, X_1=F)} = 0$

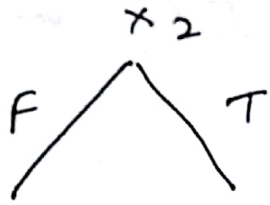
$E_{(X_2=F, X_1=T)} = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \left(\frac{1}{4}\right)\log_2\left(\frac{1}{4}\right)$

$= 0.31127 + 0.5$

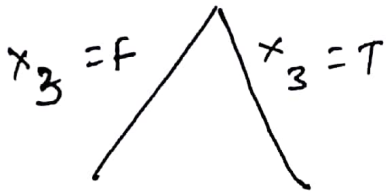
$= 0.81127$

$I.G = 0.9851 - \frac{3}{7}(0) - \frac{4}{7}(0.81127)$

$I.G = 0.52151$



Instances	1, 3, 4, 5, 8, 9, 10	2, 6, 7
class	1, 0, 0, 0, 1, 0, 1	1, 1, 1



Instances	1, 3, 5, 8, 9, 10	4,
class	1, 0, 0, 1, 0, 1	0

wkT, E_2 $(X_2 = F)$ $= \underline{0.9851}$

E_1 $(X_2 = F, X_3 = T)$ $= 0$

E_3 $(X_2 = F, X_3 = F)$ $= -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)$

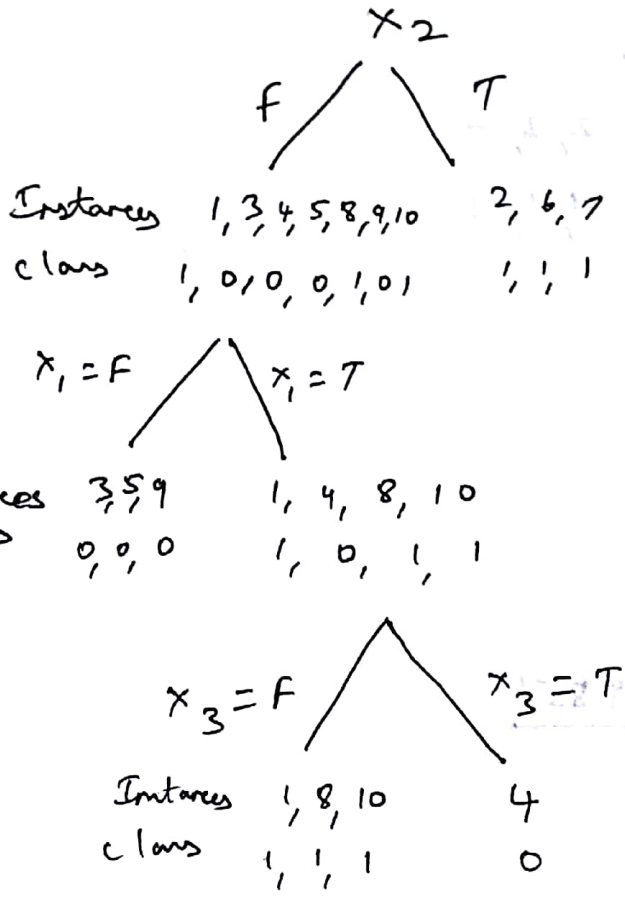
$= \frac{1}{2} + \frac{1}{2}$

$= 1$

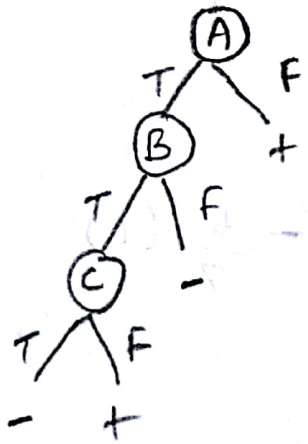
$IG = 0.9851 - \frac{1}{7}(0) - \frac{6}{7}(1)$

$= \underline{0.12795}$

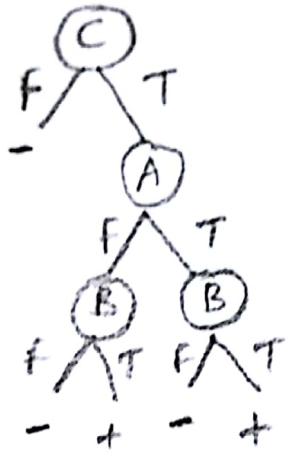
∴ x_1 has more IG, we choose it.



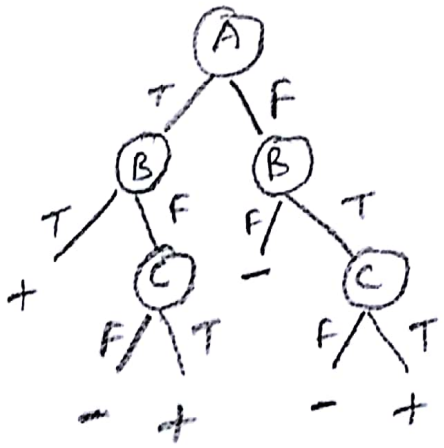
1. a) $Y = (\neg A \vee B) \wedge \neg (C \wedge A)$.



b) $Y = (A \oplus B) \wedge C = (A \wedge \neg B) \vee (\neg A \wedge B) \wedge C$



c) $Y = (A \vee B) \wedge (B \vee C) \wedge (A \vee C)$



d) $Y = (A \vee B) \wedge \neg A \wedge \neg B$