Question 1:

Bias - Variance Trade-off :

Bias variance Trade-off is a supervised machine learning algorithm to achieve low bias & low variance. In turn the algorithm should achieve good prediction performance.

For accurate prediction of the model, algorithms need a low variance & low bias. But bias & variance are related to each other. Therefore getting low bias & variance is not possible.

When it comes to supervised learning its not possible to accurately capture the data & simultaneously generalizes ~~the~~ well with the unseen dataset. Because high variance algorithm may perform well with training data, but it may lead to overfitting to noisy data.

Therefore to get the optimal model we need to find a sweet spot between bias & variance.

Question 2 :

		Predict	
		Class 1	Class 2
Actual	Class 1	50 (TP)	30 (FN)
	Class 2	40 (FP)	60 (TN)

1) Precision $= \dfrac{TP}{TP + FP}$

$= \dfrac{50}{50 + 40} = \underline{\underline{0.555}}$

2) Recall $= \dfrac{TP}{TP + FN} = \dfrac{50}{50 + 30} = \underline{\underline{0.625}}$

3) F1 score $= \dfrac{2}{\dfrac{1}{Precision} + \dfrac{1}{Recall}} = \dfrac{2\,(Precision \times Recall)}{Precision + Recall}$

$= \dfrac{2\,(0.555 \times 0.625)}{0.555 + 0.625} = \dfrac{2.36}{1.18} = \underline{\underline{2}} = 0.585$

Question 3 :-

In this data set we have 4 features & 2 classes

$$E(S) = -\sum_{i=1}^{n} P_i \log(P_i)$$

$$E(6+, 4-) = -\frac{6}{10}\log\left(\frac{6}{10}\right) - \frac{4}{10}\log\left(\frac{4}{10}\right)$$

$$= 0.443 + 0.528$$

$$= \underline{0.971}$$

Lets get the entropy of 4 features ~~by taking~~ & ~~get~~ the gain

~~Outlook~~ ~~Growth~~

Lets get Gain

$$E(\text{outlook}, \text{sunny}) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$$

$$= 0.811$$

$$E(\text{outlook}, \text{overcast}) = 0$$

$$E(\text{outlook}, \text{Rain}) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$$

$$= 0.811$$

So $\text{Gain}(S, \text{outlook}) = \text{Entropy}(S) - \sum \frac{|S_v|}{S}\,\text{Entropy}(S_v)$

$$= 0.971 - 4/10\,(0.811) - 0 - 4/10\,(0.811)$$

$$= \underline{0.322}$$

Lets directly get $\sum \frac{|S_v|}{S}$ Entropy (S_v) of Temperature.

$$= \frac{3}{10}\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right) + \frac{3}{10}\left(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right)$$

$$+ \frac{4}{10}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right)$$

$$= 0.875$$

For Humidity

$$= 5/10\left(-2/5\log_2 2/5 - 3/5\log_2 3/5\right) + 5/10\left(-4/5\log_2\frac{4}{5} - 1/5\log_2 1/5\right)$$

$$= 0.846$$

For Wind

$$= 7/10\left(-5/7\log_2 5/7 - 2/7\log_2 2/7\right) + 3/10\left(-1/3\log_2 1/3 - 2/3\log_2 2/3\right)$$

$$= 0.879$$

Gain of Temp, Humi & wind

$G(S, \text{Temp}) = 0.096$

$G(S, \text{Hum}) = 0.125$

$G(S, \text{Wind}) = 0.092$

Outlook has the highest gain.

Next best attribute for split

$E(\text{outlook sunny}) = 0.811$

~~$E(\text{sunny, Temp})$~~ $= \sum \frac{|S_v|}{S}$ entropy S_v for the attributes

For Temp → $\frac{2}{4}\left(-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right) + \frac{1}{4}\left(-\frac{0}{1}\log_2\left(\frac{0}{1}\right)\right.$

$\left. - \frac{1}{1}\log_2\left(\frac{1}{1}\right)\right) = 0$

For Humidity $= 0$

For wind $= \frac{3}{4}\left(-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right)$

$= 0.689$

$\text{Gain}_s(S, \text{Temp}) = 0.811 - 0 = 0.811$

$\text{Gain}_s(S, \text{Hum}) = 0.811$

$\text{Gain}(S, \text{wind}) = 0.122$

~~So~~ Here both Temp & Hum have same gain. Therefore choose any one of them.

(choosing Temp)

out look == Rain

$E(\text{outlook, Rain}) = 0.811$

$\sum \frac{|S_v|}{S}$ entropy S_v for all the attribute

For Temp = $\cancel{\frac{2}{4}}$ - 1

For Hum = $\frac{3}{4}\left(-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) = 0.689.$

For wind = 0

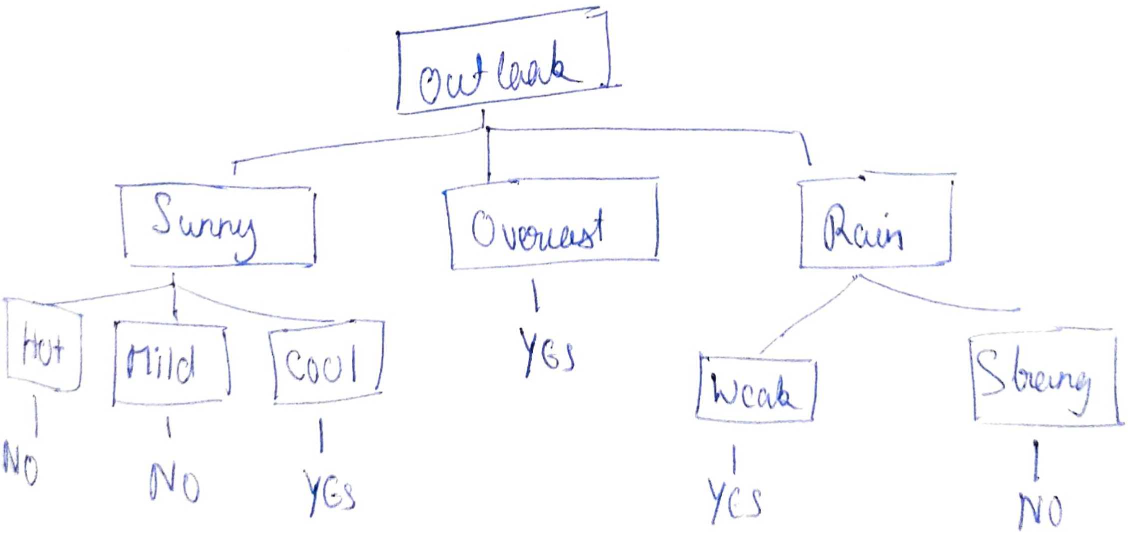
Here

$Gain_R(S, Temp) = 0.811 - 1 = -0.189$

$Gain(S, Hum) = 0.122$

$Gain(S, wind) = 0.811$

We have to split wind in the tree



Question 4 :- Naive Bays method :-

Bays formula $\rightarrow$ $P(w|x) = \dfrac{P(x/w)\,P(w)}{P(x)}$

$$M_i(x) \propto \prod_{i=1} \hat{P}(w_i \,|\, d_{ij_i}(x)=1)$$

Here

$P(w_1/d_{11}(x=1)) = \frac{40}{70}$

$P(W_2/d_{12}(x=1)) = \frac{30}{70}$

$P(W_2/d_{21}(x=1)) = \frac{20}{40}$

$P(w_2/d_{22}(x=1)) = \frac{20}{40}$

$P(W_3/d_{31}(x=1)) = 0/10$

$P(W_3/d_{32}(x=1)) = 10/10.$

For class 1

$$\frac{40}{70} \times \frac{20}{40} \times \frac{0}{10} = 0$$

For class 2

$$\frac{30}{70} \times \frac{20}{40} \times \frac{10}{10} = 0.214$$