No: 22120931 (NUR HIDAYAH BINTI AHMAD SHAFII)

1. a. ii) $P(\text{Win} \mid \text{Pasta}) = 0.4$

$P(\text{Win} \mid \text{Sushi}) = 0.25$

$P(\text{Win} \mid \text{Salad}) = 0.1$

$P(\text{Pasta}) = P(\text{Sushi}) = P(\text{Salad}) = \frac{1}{3}$

i) $P(\text{Pasta} \mid \text{Win}) = \dfrac{P(\text{Win} \mid \text{Pasta}) \cdot P(\text{Pasta})}{P(\text{Win})}$

$P(\text{Win}) = P(\text{Win} \mid \text{Pasta}) \cdot P(\text{Pasta}) + P(\text{Win} \mid \text{Sushi}) \cdot P(\text{Sushi}) + P(\text{Win} \mid \text{Salad}) \cdot P(\text{Salad})$

$P(\text{Win}) = 0.4\left(\frac{1}{3}\right) + 0.25\left(\frac{1}{3}\right) + 0.1\left(\frac{1}{3}\right)$

$= \dfrac{0.75}{3}$

$= 0.25$

$P(\text{Pasta} \mid \text{Win}) = \dfrac{0.4\left(\frac{1}{3}\right)}{0.25}$

$= 0.5333 \approx 53.33\%$

$\therefore$ the probability that he win based on pasta is approximately 53.33%

No:

1.a.ii) $P(\text{Sushi}) = 0.8$

$P(\text{Salad}) = 0.2$

$P(\text{Pasta}) = 0$

$P(\text{Win} \mid \text{Sushi}) = 0.25$

$P(\text{Win} \mid \text{Salad}) = 0.1$

$$P(\text{Win}) = P(\text{Win} \mid \text{Sushi}) \cdot P(\text{Sushi}) + P(\text{Win} \mid \text{Salad}) \cdot P(\text{Salad})$$
$$= 0.25(0.8) + 0.1(0.2)$$
$$= 0.22$$

$$P(\text{Salad} \mid \text{Win}) = \frac{P(\text{Win} \mid \text{Salad}) \cdot P(\text{Salad})}{P(\text{Win})}$$
$$= \frac{0.1(0.2)}{0.22}$$
$$\approx 0.0909$$
$$\approx 9.09\%$$

$\therefore$ the probability that Liam won because he made salad is approximately 9.09 %.

No:

1.b.i) Logistic Regression.
This is because it involves binary classification of $Y \in \{0,1\}$
It can regress on linearly separable data.

ii) Yes. Logistic regression using the sigmoid function will allow linear decision boundary in the feature space.
A linear decision boundary ($x_1 + x_2 = 0.5$) can separate the positive ~~values~~ points (0,1), (1,1), (1,0) from negative points (0,0)

iii) No, there ~~will~~ no longer perfect logistic regression classifier.
This is because the new data points cannot be linearly separated.
~~For example~~, Any linear decision boundary would misclassify at least one point.

1.b.i) Logistic Regression. This is because the problem involves binary classification and the model can regress on linearly separable data. Also, the output possibilities is specifically designed for classification tasks.

ii) No. This is because logistic regression can only separate data points with straight line. The current arrangement of points is not linearly separable. Therefore, no straight line can perfectly separate the positive and negative samples

iii) Yes. This because the data will become linearly separable. A straight line can be drawn to perfectly separate positive and negative samples. The model will be able to find perfect classifier in this modified scenario.

No:

2.a) ~~2-b~~ ~~2-b~~

Total = 6

Yes (Hired) = 3

No (Not Hired) = 3

$P(Yes) = \frac{3}{6} = 0.5$

$P(No) = \frac{3}{6} = 0.5$

$\text{Entropy}(S) = -[0.5 \log_2(0.5) + 0.5 \log_2(0.5)]$

$= 1$

∴ the entropy of the dataset is 1

b) High : (2 Yes, 0 No)

Medium : (1 Yes, 1 No)

Low : (0 Yes, 2 No)

$\text{Entropy}(\text{High}) = -\left[\frac{2}{2} \log_2\left(\frac{2}{2}\right) + \frac{0}{2} \log_2\left(\frac{0}{2}\right)\right] = 0$

$\text{Entropy}(\text{Medium}) = -\left[\frac{1}{2} \log_2\left(\frac{1}{2}\right) + \frac{1}{2} \log_2\left(\frac{1}{2}\right)\right] = 1$

$\text{Entropy}(\text{Low}) = -\left[\frac{0}{2} \log_2\left(\frac{0}{2}\right) + \frac{1}{2} \log_2\left(\frac{2}{2}\right)\right] = 0$

$\text{Entropy}(\text{Qualifications}) = \frac{2}{6}(0) + \frac{2}{6}(1) + \frac{2}{6}(0)$

$= \frac{2}{6} = 0.333$

No:

2.b. For interview score

High : (2 Yes, 1 No)

Low (1 Yes, 2 No)

$$\text{Entropy (High)} = -\left[\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] = 0.918$$

$$\text{Entropy (Low)} = -\left[\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\left(\log_2\left(\frac{2}{3}\right)\right]\right. = 0.918$$

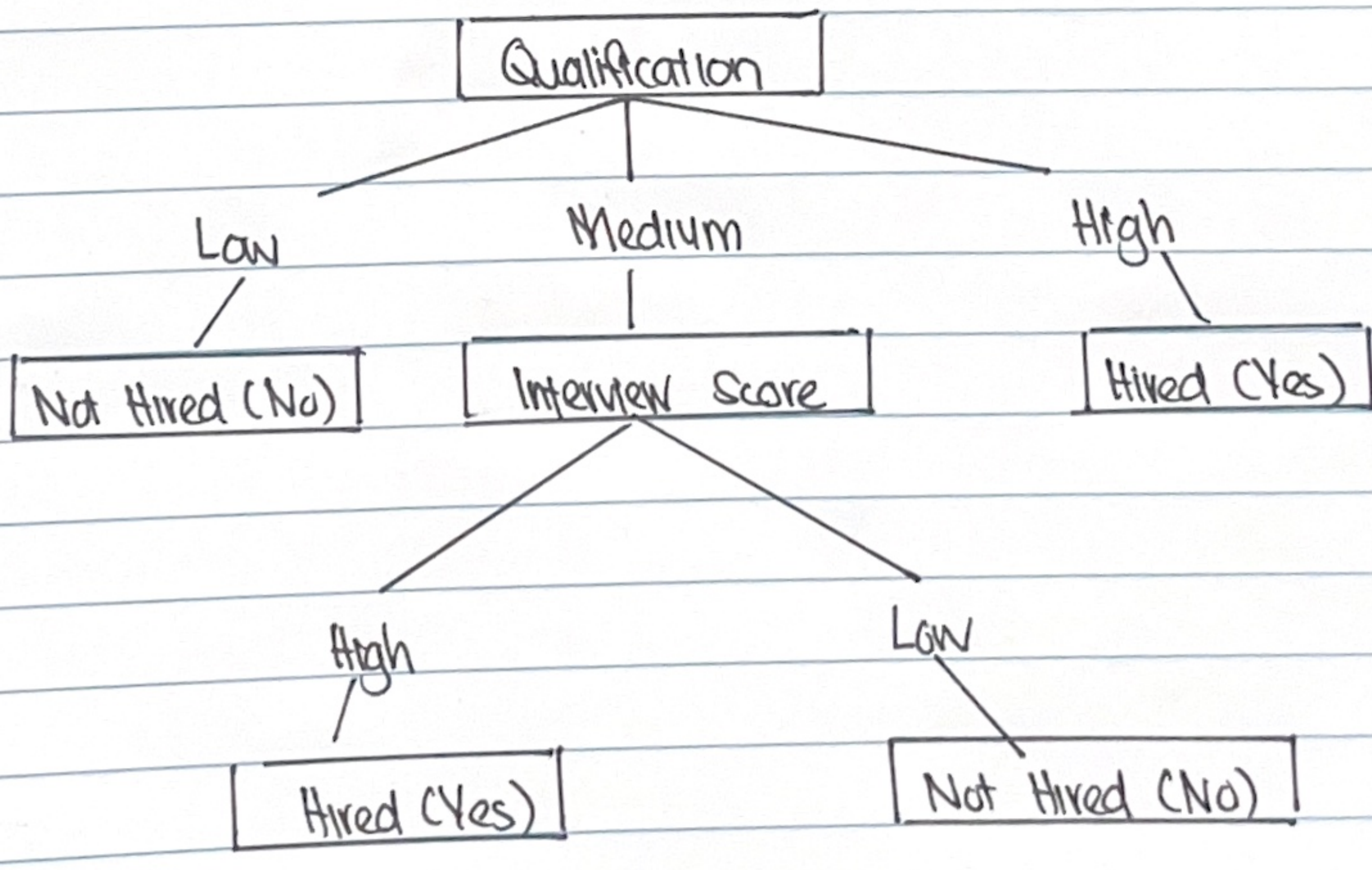
$$\text{Entropy (Interview Score)} = \frac{3}{6}(0.918) + \frac{3}{6}(0.918) = 0.918$$

2.c) Gain (S, Qualifications) = 1 - 0.333 = 0.667

Gain (S, Interview Score) = 1 - 0.918 = 0.082

∴ the feature with the highest information gain is qualification.

2.d)



No:

3.

	A	B	C	D
A	0	1	4	5
B	1	0	2	6
C	4	2	0	3
D	5	6	3	0

min distance, $d(A,B) = \min\{d(x,y) : x \in A, y \in B\}$

C_1 : $\{A\}, \{B\}, \{C\}, \{D\}$

From the table, the ~~[illegible]~~ minimum distance is the distance between the clusters $\{A\}$ and $\{B\}$ which is 1

$d[(AB), C] = \min[d(A,C), d(B,C)\}$
$= \min(4, 2)$
$= 2$

$d[(A,B), D] = \min\{d(A,D), d(B,D)\}$
$= \min(5, 6)$
$= 5$

~~A~~ A, B	A, B	C	D
A, B	0	2	5
C	2	0	3
D	5	3	0

∴ ~~the minimum distance between (A,B) and C~~

∴ the minimum distance is the distance between the cluster $\{A, B\}$ and $\{C\}$ which is 2.

$d[(A,B,C), E] = \min\{d(A,D), d(B,D), d(C,D)\}$
$= \min\{5, 6, 3\}$
$= 3$

No:

3-

	A,B,C	D
A,B,C	0	3
D	3	0

∴ the minimum distance is the distance between cluster {A,B,C} and {D} which is 3. It will perform new cluster of {A,B,C,D}