1.c)

Record	Study hours	attendance	Prepration	passed exam
1	High	High	Good	yes
2	High	High	Average	yes
3	Medium	High	Good	yes
4	Low	High	Good	yes
5	Low	Low	Poor	NO
6	Medium	Low	Average	No
7	Medium	Low	Poor	No
8	Low	High	Average	No
9	High	Low	Good	yes
10	High	Low	Poor	No
11	Medium	High	Average	yes
12	Medium	High	Poor	No
13	Low	High	Poor	No
14	High	High	Poor	No
15	High	Low	Average	yes

Training Set: records 1–10

Test Set: records 11–15

$$Info(P) = -P_{yes} \cdot \log_2(P_{yes}) - P_{No} \cdot \log_2(P_{no})$$

$$P_{yes} = P_{NO} = 0.5$$

$$Info(P) = -0.5 \cdot \log_2(0.5) - 0.5 \cdot \log_2(0.5) = 1$$

$$Info(P) = 1$$

1.ב)

$$Gain(P, StudyHours) = Info(P) - Info_{StudyHours}(P)$$

$$Info_{StudyHours}(P) = \frac{|P_{High}|}{10} \cdot info(P_{High}) + \frac{|P_{Medium}|}{10} \cdot info(P_{Medium}) + \frac{|P_{Low}|}{10} \cdot info(Low) =$$

$$= \frac{4}{10} \cdot info(P_{High}) + \frac{3}{10} \cdot info(P_{Medium}) + \frac{3}{10} \cdot info(P_{Low})$$

$$Info(P_{High}) = -P_{yes/High} \cdot \log_2(P_{yes/High}) - P_{No/High} \cdot \log_2(P_{No/High})$$

$$Info(P_{High}) = -\frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) = \underline{0.811}$$

$$Info(P_{Medium}) = -P_{yes/Medium} \cdot \log_2(P_{yes/Medium}) - P_{No/Medium} \cdot \log_2(P_{No/Medium})$$

$$Info(P_{Medium}) = -\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) = \underline{0.918}$$

$$Info(P_{Low}) = -P_{yes/Low} \cdot \log_2(P_{yes/Low}) - P_{No/Low} \cdot \log_2(P_{No/Low})$$

$$Info(P_{Low}) = -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) = \underline{0.918}$$

$$Info_{StudyHours}(P) = \frac{4}{10} \cdot 0.811 + \frac{3}{10} \cdot 0.918 + \frac{3}{10} \cdot 0.918 = \underline{0.8752}$$

$$Gain(P, StudyHours) = 1 - 0.8752 = 0.1248$$

$$Gain(P, StudyHours) = \underline{0.1248}$$

$$Gain(P, Attendance) = Info(P) - Info_{Attendance}(P)$$

$$Info_{Attendance}(P) = \frac{|P_{High}|}{10} \cdot Info(P_{High}) + \frac{|Low|}{10} \cdot Info(P_{Low})$$

$$Info_{Attendance}(P) = \frac{1}{2} \cdot Info(P_{High}) + \frac{1}{2} Info(P_{Low})$$

1. K)

$Info(P_{High}) = -P_{yes/High} \cdot \log_2(P_{yes/High}) - P_{No/High} \cdot \log_2(P_{No/High})$

$Info(P_{High}) = -\frac{4}{5} \cdot \log_2(\frac{4}{5}) - \frac{1}{5} \cdot \log_2(\frac{1}{5}) = 0.721$

$Info(P_{Low}) = -P_{yes/Low} \cdot \log_2(P_{yes/Low}) - P_{No/Low} \cdot \log_2(P_{No/Low})$

$Info(P_{Low}) = -\frac{1}{5} \cdot \log_2(\frac{1}{5}) - \frac{4}{5} \cdot \log_2(\frac{4}{5}) = 0.721$

$Info_{Attendance}(P) = \frac{1}{2} \cdot 0.721 + \frac{1}{2} \cdot 0.721 = 0.721$

$Gain(P, Attendance) = 1 - 0.721 = 0.279$

$Gain(P, Attendance) = \underline{0.279}$

$Gain(P, preparation) = Info(P) - Info_{prepration}(P)$

$Info_{prepration}(P) = \frac{|P_{Good}|}{10} \cdot info(P_{Good}) + \frac{|P_{Average}|}{10} \cdot info(P_{Average}) +$

$+ \frac{|P_{Poor}|}{10} \cdot info(P_{Poor})$

$Info_{prepratin}(P) = \frac{4}{10} \cdot info(P_{Good}) + \frac{3}{10} \cdot info(P_{Average}) + \frac{3}{10} \cdot info(P_{Poor})$

$Info(P_{Good}) = -P_{yes/Good} \cdot \log_2(P_{yes/Good}) - P_{No/Good} \cdot \log_2(P_{No/good})$

$Info(P_{Good}) = -1 \cdot \log_2(1) - 0 = 0$

$Info(P_{Average}) = -P_{yes/Average} \cdot \log_2(P_{yes/Average}) - P_{No/Average} \cdot \log_2(P_{No/Average})$

$Info(P_{Average}) = -\frac{1}{3} \cdot \log_2(\frac{1}{3}) - \frac{2}{3} \cdot \log_2(\frac{2}{3}) = 0.918$

$Info(P_{Poor}) = -P_{yes/Poor} \cdot \log_2(P_{yes/Poor}) - P_{No/Poor} \cdot \log_2(P_{No/poor})$

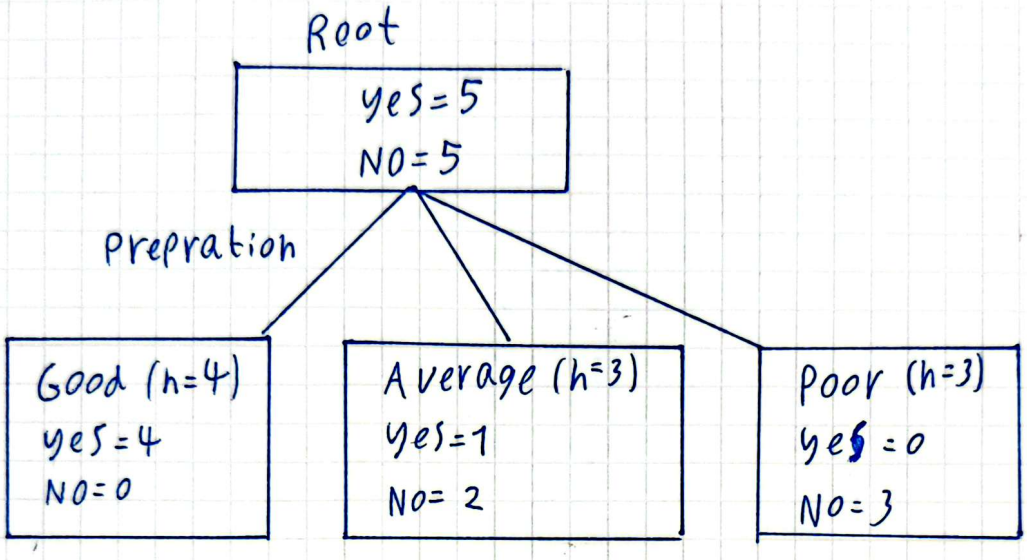
$Info(P_{Poor}) = -0 - 1 \cdot \log_2(1) = 0$

$Info_{preparation}(P) = 0 + \frac{3}{10} \cdot 0.918 + 0 = 0.2754$

$Info_{preparation}(P) = 0.2754$

$Gain(P, Preparation) = \underline{0.7246}$

Attribute	Information Gain
Study Hours	0.124
Attendance	0.279
Prepration	0.724



Average:

Record	Study Hours	Attendance	Passed Exam
1	High	High	yes
2	Medium	Low	NO
3	Low	High	NO

$Info(P_{Average}) = 0.918$

$Gain(P_{Average}, StudyHours) = Info(Average) - Info_{StudyHours}(P_{Average})$

$Info_{StudyHours}(P_{Average}) = \frac{|P_{High}|}{3} \cdot Info(P_{High}) + \frac{|P_{Medium}|}{3} + Info(P_{Medium}) + \frac{|P_{Low}|}{3} \cdot Info(P_{Low})$

$Info_{StudyHours}(P_{Average}) = \frac{1}{3} Info(P_{High}) + \frac{1}{3} Info(P_{Medium}) + \frac{1}{3} \cdot Info(P_{Low})$

$Info(P_{High}) = -P_{yes/High}\log_2(P_{yes/High}) - P_{NO/High}\cdot\log_2(P_{NO/High})$

$Info(P_{High}) = 1\cdot\log_2(1) - 0 = 0$

$Info(P_{Medium}) = -P_{yes/Medium}\cdot\log_2(P_{yes/Medium}) - P_{NO/Medium}\cdot\log_2(P_{NO/Medium})$

$Info(P_{Medium}) = 0 - 1\cdot\log_2(1) = 0$

$Info(P_{Low}) = -P_{yes/Low}\log_2(P_{yes/Low}) - P_{NO/Low}\cdot\log_2(P_{NO/Low})$

$Info(P_{Low}) = 0 - 1\cdot\log_2(1) = 0$

$Info_{StudyHours}(P_{Average}) = 0$

$Gain(P_{Average}, StudyHours) = 0.918 - 0 = 0.918$

$Info_{Attendance}(P_{Average}) = \frac{|P_{High}|}{3}\cdot Info(P_{High}) + \frac{|P_{Low}|}{3}\cdot Info(P_{Low})$

$Gain(P_{Average}, Attendance) = Info(P_{Average}) - Info_{Attendance}(P_{Average})$

$Info(P_{High}) = -P_{yes/High}\cdot\log_2(P_{yes/High}) - P_{NO/High}\cdot\log_2(P_{NO/High})$

$Info(P_{High}) = -\frac{1}{2}\cdot\log_2(\frac{1}{2}) - \frac{1}{2}\cdot\log_2(\frac{1}{2}) = 1$

$Info(P_{Low}) = 0 - 1\cdot\log_2(1) = 0$

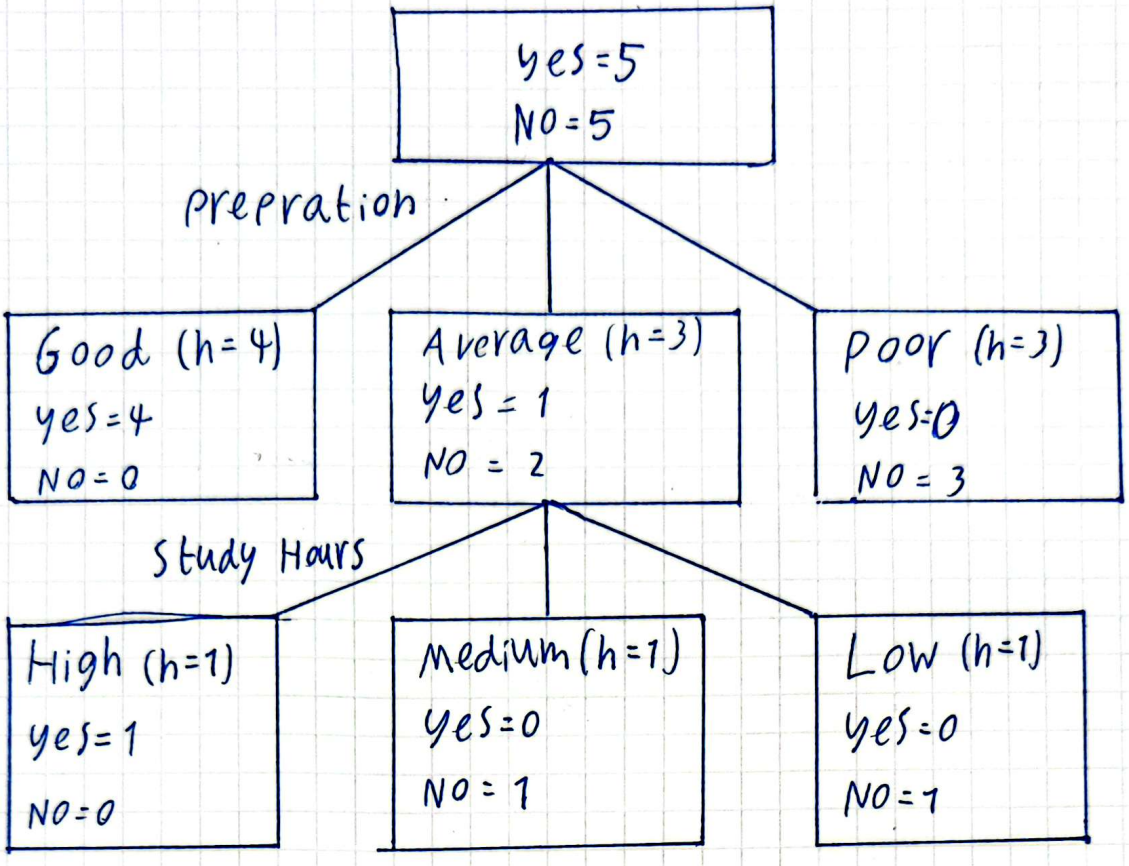
$Info_{Attendance}(P_{Average}) = \frac{2}{3}\cdot Info(P_{High}) + \frac{1}{3}\cdot Info(P_{Low})$

$Info_{Attendance}(P_{Average}) = \frac{2}{3}$

$Gain(P_{Average}, Attendance) = 0.918 - \frac{2}{3} = 0.251$

Attribute	Information Gain
Study Hours	0.918
Attendance	0.251

1.c)



1.2) $\frac{10}{10} = 1$

Accuracy training = 100%

1.d) $\frac{4}{5} = 0.8$

Accuracy test = 80%