(d) Class labels: Yes → 9, No → 5

$$Entropy(D) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.9403$$

- $Gain(D, type) = 0.9403 - \left(\frac{3}{7}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{2}{7}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right)\right.$

math: 6 (Yes→3, no→3)
language: 4 (yes→3, no→1) $\left. + \frac{2}{7}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right)\right) = 0.0481411$
density: 4 (yes→3, no→1)

- $Gain(D, difficulty) = 0.9403 - \left(\frac{2}{7}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{3}{7}\left(\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right)\right.$

hard: 4 (Yes→2, no→2)
medium: 6 (yes→4, no→2) $\left. + \frac{2}{7}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right)\right) = 0.0292366$
easy: 4 (yes→3, no→1)

~ $Gain(D, learned\_before) = 0.9403 - \left(\frac{4}{7}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right) + \frac{3}{7}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right)\right)$

Yes: 8 (yes→6, no→2) $= 0.0481411$
no: 6 (yes→3, no→3)

- $Gain(D, completeness) = 0.9403 - \left(\frac{9}{14}\left(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right) + \frac{5}{14}\left(-\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}\right)\right)$

poor: 9 (yes→6, no→3) $= 0.0031989$
good: 5 (yes→3, no→2)

Choose either type or learned_before.

(e) using same information gain value from (d),

$$SplitInfo_{type} = -\frac{3}{7}\log_2\frac{3}{7} - \frac{2}{7}\log_2\frac{2}{7} - \frac{2}{7}\log_2\frac{2}{7} = 1.5567$$

$$SplitInfo_{difficulty} = -\frac{2}{7}\log_2\frac{2}{7} - \frac{3}{7}\log_2\frac{3}{7} - \frac{2}{7}\log_2\frac{2}{7} = 1.5567$$

$$SplitInfo_{learned\_before} = -\frac{4}{7}\log\frac{4}{7} - \frac{3}{7}\log\frac{3}{7} = 0.98523$$

$$SplitInfo_{completeness} = -\frac{9}{14}\log\frac{9}{14} - \frac{5}{14}\log\frac{5}{14} = 0.94029$$

$GainRatio(type) = 0.0481411 / 1.5567 = 0.03093$

$GainRatio(difficulty) = 0.0292366 / 1.5567 = 0.01878$

$GainRatio(learned\_before) = 0.0481411 / 0.98523 = 0.04886$

$GainRatio(completeness) = 0.0031989 / 0.94029 = 0.0034020$

Choose learned_before.

(f) ~ $Gini(math) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$, $Gini(language) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375$,

$Gini(language) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375$

$Gini(type) = \frac{3}{7}(0.5) + \frac{2}{7}(0.375) + \frac{2}{7}(0.375) = 0.42857$

~ $Gini(hard) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$, $Gini(medium) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = \frac{4}{9}$

$Gini(easy) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375$

$Gini(difficulty) = \frac{2}{7} \times 0.5 + \frac{3}{7} \times \frac{4}{9} + \frac{2}{7} \times 0.375 = 0.44048$

~ $Gini(yes) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375$, $Gini(no) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$

$Gini(learned\_before) = \frac{4}{7} \times 0.375 + \frac{3}{7} \times 0.5 = 0.42857$

~ $Gini(poor) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = \frac{4}{9}$, $Gini(good) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = 0.48$

$Gini(completeness) = \frac{9}{14}\left(\frac{4}{9}\right) + \frac{5}{14}(0.48) = 0.45714$

Choose either type or learned_before.