

(1)

classmate

Date _____

Page _____

$$\text{Entropy}(S) = -\frac{7}{20}$$

$$= -P(\text{poor}) \log_2(P(\text{poor})) - P(\text{average}) \log_2(P(\text{average}))$$

$$- P(\text{excellent}) \log_2(P(\text{excellent}))$$

$$= -\frac{7}{20} \log_2\left(\frac{7}{20}\right) - \frac{7}{20} \log_2\left(\frac{7}{20}\right) - \frac{6}{20} \log_2\left(\frac{6}{20}\right)$$

$$= 1.581$$

Formulae

$$\text{Entropy} = \frac{-p}{p+n} \log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n} \log_2\left(\frac{n}{p+n}\right)$$

average information

$$I(\text{attribute}) = \sum \frac{p_i + n_i}{p+n} \text{Entropy}(A)$$

$$\text{Gain} = \text{Entropy}(S) - I(\text{Attribute})$$

'Academics'

	poor	avg	excellent	entropy
good bad fail	2	0	1	0.9182
good pass	5	6	0	0.994
very good distinction	0	1	5	0.65

$$\text{Entropy (academics = 'fail')} = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - 0 - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$$

$$E(\text{academics = 'good'}) = -\frac{5}{11} \log_2\left(\frac{5}{11}\right) - \frac{6}{11} \log_2\left(\frac{6}{11}\right) - 0 = 0.994$$

$$E(\text{academics = 'very good'}) = 0 - \frac{1}{6} \log_2\left(\frac{1}{6}\right) - \frac{5}{6} \log_2\left(\frac{5}{6}\right) = 0.65$$

$$I(\text{academics}) = \left(\frac{2+0+1}{20} * 0.918\right) + \left(\frac{5+6+0}{20} * 0.994\right) + \left(\frac{0+1+5}{20} * 0.65\right) = 0.8794$$

$$\text{Gain} = \text{Entropy (academics)} - I(\text{academics}) = 1.581 - 0.8794$$

$$\text{Gain (academics)} = 0.7016$$

'speaking'

	poor	average	excellent	entropy
bad ^{hustant}	6	2	2	1.371
good fair ^{fluent}	1	5	0	0.65
very good ^{fluent}	0	0	4	0

$$E(\text{speaking} = \text{'bad'}) = \frac{-6}{10} \log_2\left(\frac{6}{10}\right) - \frac{2}{10} \log_2\left(\frac{2}{10}\right) - \frac{2}{10} \log_2\left(\frac{2}{10}\right)$$

$$= 1.371$$

$$E(\text{speaking} = \text{'good'}) = \frac{-1}{6} \log_2\left(\frac{1}{6}\right) - \frac{5}{6} \log_2\left(\frac{5}{6}\right) - 0$$

$$= 0.65$$

$$E(\text{speaking} = \text{'very good'}) = 0 - 0 - \frac{4}{4} \log_2\left(\frac{4}{4}\right) = 0$$

$$I(\text{speaking}) = \left[\left(\frac{6+2+2}{20} \right) * 1.371 \right] + \left[\left(\frac{1+5}{20} \right) * 0.65 \right]$$

$$+ \left[\left(\frac{0+0+4}{20} \right) * 0 \right]$$

$$= 0.8805$$

$$\text{Gain}(\text{speaking}) = E(S) - I(\text{speaking})$$

$$= 1.581 - 0.8805$$

$$= 0.7005$$

'creative'

	poor	average	excellent	entropy
bad ^{low}	4	1	0	0.722
good ^{medium}	3	4	3	1.571
very good ^{high}	0	2	3	0.971

$$\text{Entropy (creative = 'bad' ^{low})} = \frac{-4}{5} \log_2\left(\frac{4}{5}\right) - \frac{1}{5} \log_2\left(\frac{1}{5}\right) - 0$$

$$= 0.722$$

$$\text{Entropy (creative = 'good' ^{medium})} = \frac{-3}{10} \log_2\left(\frac{3}{10}\right) - \frac{4}{10} \log_2\left(\frac{4}{10}\right) - \frac{3}{10} \log_2\left(\frac{3}{10}\right)$$

$$= 1.571$$

$$\text{Entropy (creative = 'very good' ^{high})} = \frac{-2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) - 0$$

$$= 0.971$$

$$I(\text{creative}) = \left[\left(\frac{4+1}{20} \right) * 0.722 \right] + \left[\left(\frac{3+4+3}{20} \right) * 1.571 \right] + \left[\left(\frac{0+2+3}{20} \right) * 0.971 \right]$$

$$= 1.20875$$

$$\text{Gain (creative)} = E(S) - I(\text{creative})$$

$$= 1.581 - 1.20875$$

$$= 0.37225$$

'sports'

	poor	average	excellent	entropy
bad	7	3	0	0.881
good	0	3	3	1
very good	0	1	3	0.811

$$E(\text{sports} = \text{bad}) = -\frac{7}{10} \log_2\left(\frac{7}{10}\right) - \frac{3}{10} \log_2\left(\frac{3}{10}\right) - 0$$

$$= 0.881$$

$$E(\text{sports} = \text{good}) = -\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) - 0$$

$$= 1$$

$$E(\text{sports} = \text{very good}) = -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) - 0$$

$$= 0.811$$

$$I(\text{sports}) = \left[\left(\frac{7+3+0}{20} \right) * 0.881 \right] + \left[\left(\frac{0+3+3}{20} \right) * 1 \right]$$

$$+ \left[\left(\frac{1+3+0}{20} \right) * 0.811 \right]$$

$$= 0.9027$$

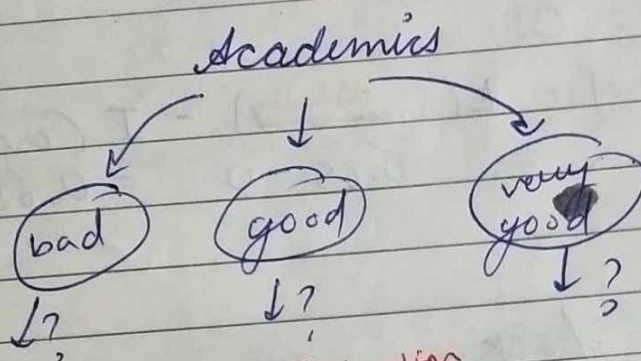
$$\text{Gain}(\text{sports}) = E(S) - I(\text{sports})$$

$$= 1.581 - 0.9027$$

$$= 0.6783$$

Attributes	Gain
Academics	0.7016
Speaking	0.7005
Creative	0.3722
Sports	0.6783

∴ Root Node = Academics



Academics (~~very good~~ ^{distinction})

total data points = 6

$$E(S_{\text{very good}}) = \frac{5}{6} \log_2 \left(\frac{5}{6} \right) - \frac{1}{6} \log_2 \left(\frac{1}{6} \right) = 0.65$$

Speaking	poor	avg	excellent	entropy
bad ^{hesitant}	0	2	1	1
good ^{fair}	0	0	0	0
very good ^{fluent}	0	0	4	0

$$E(\text{speaking} = \text{bad}) = \frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) = 1$$

$$E(\text{speaking} = \text{good}) = 0$$

$$E(\text{speaking} = \text{very good}) = \frac{4}{4} \log_2 \left(\frac{4}{4} \right) = 0$$

$$I(\text{speaking}) = \left(\frac{2}{6} * 1 \right) = 0.333$$

$$\text{Gain} = E(S_{\text{very good}}) - I(\text{speaking}) = 0.65 - 0.33 = 0.32$$

Creative	poor	avg	excellent	entropy
bad ^{low}	0	1	0	0
good ^{med}	0	0	3	0
very good ^{high}	0	0	2	0

$$E(\text{bad}) = 0$$

$$E(\text{good}) = 0$$

$$E(\text{very good}) = 0$$

$$I(\text{creative}) = 0$$

$$\text{Gain}(\text{creative}) = E(S_{\text{very good}}^{\text{distinction}}) - I(\text{creative})$$

$$= 0.65 - 0 = 0.65$$

Speeds

	poor	avg	excellent	entropy
bad	0	0	0	0
good	0	1	3	0.811
very good	0	0	2	0

$$E(\text{bad}) = 0$$

$$E(\text{good}) = \frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) = 0.811$$

$$E(\text{very good}) = 0$$

$$I(\text{speeds}) = \frac{4}{6} \times 0.811 = 0.5406$$

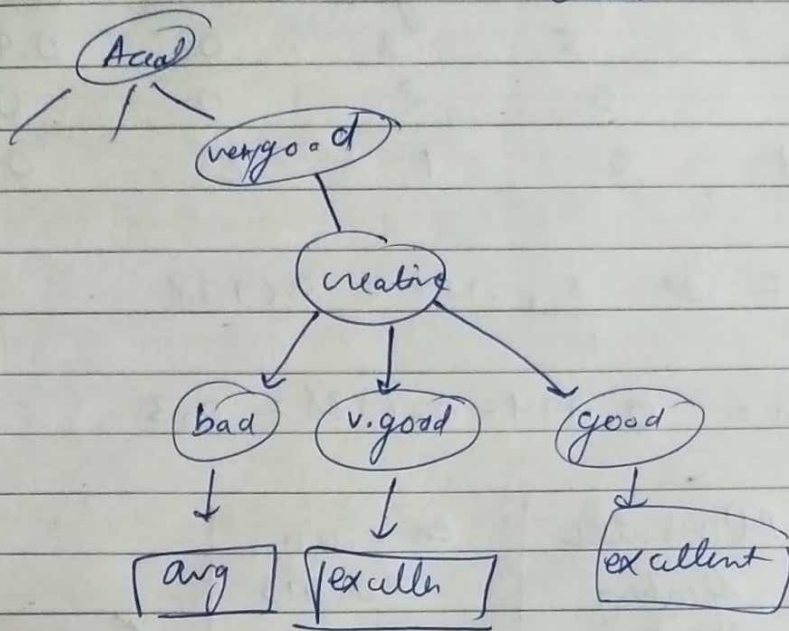
$$\text{Gain}(\text{speeds}) = E(S_{\text{very good}}^{\text{distinction}}) - I(\text{speeds})$$

$$= 0.65 - 0.5406 = 0.1094$$

Attribute	Gain	
Speaking	0.32	
Creative	0.65	✓
Speeds	0.1094	

∴ Next Node under 'very good' of Academics

will be creative



Academics → ^{poor}good data points = 11
 $E(S_{\text{good}}^{\text{poor}}) = 0.994$

Speaking

	poor	avg	excellent	entropy
bad ^{hesitant}	4	1	0	0.722
good ^{fair}	1	5	0	0.65
v. good ^{fluent}	0	0	0	0

$$I = \left[\left(\frac{5}{11} \right) * 0.722 \right] + \left[\left(\frac{6}{11} \right) * 0.65 \right]$$

$$= 0.6827$$

gain = 0.3113

Creative

	poor	avg	excellent	entropy
bad ^{low}	3	0	0	0
good ^{med}	2	4	0	0.918
very good ^{high}	0	2	0	0

$$I = \frac{6}{11} * 0.918 = 0.5$$

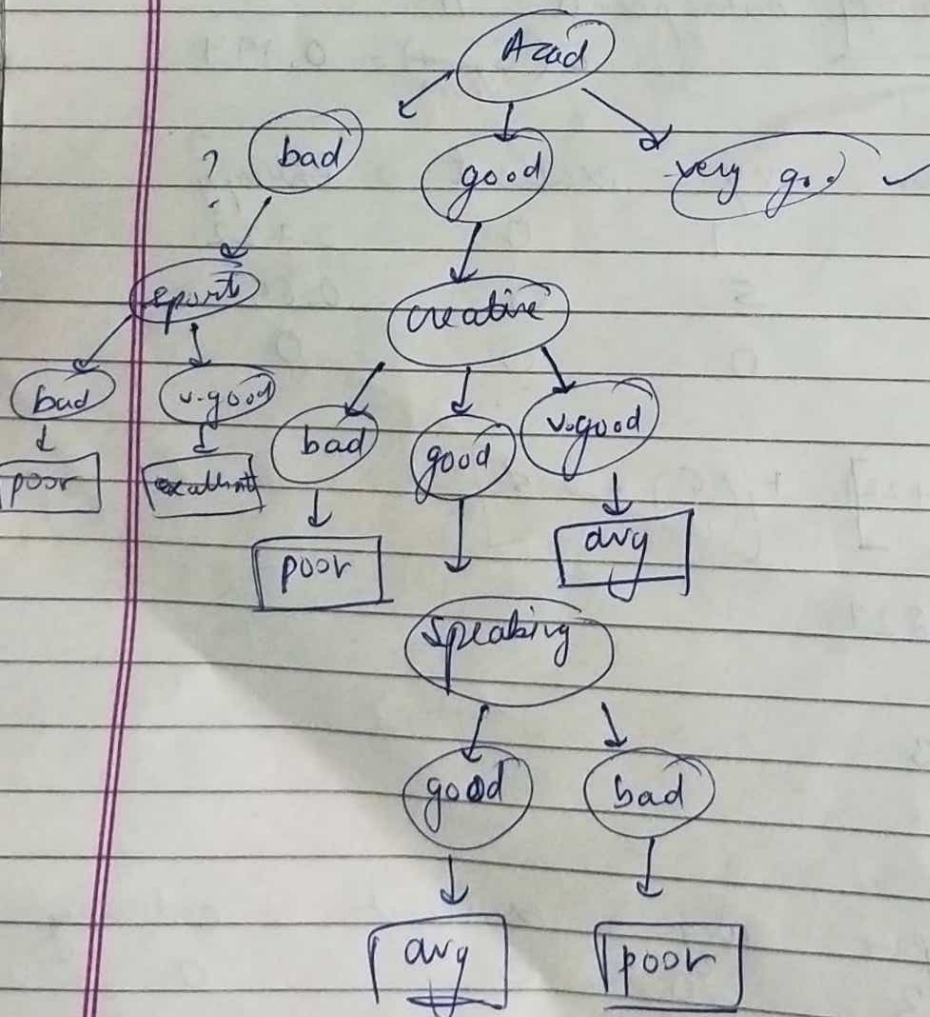
Gain = 0.494

<u>Sports</u>	poor	avg	excellent	entropy
bad	5	3	0	0.954
good	0	2	0	0
v good	0	1	0	0

$$I = \frac{8}{11} \times 0.954 = 0.6938$$

$$\text{Gain} = 0.994 - 0.6938 = 0.3$$

<u>Attributes</u>	<u>en gain</u>
Speaking	0.3113
Creative	0.494
Sports	0.30



Academics

Academics ^{poor} ~~good~~ creative ~~good~~ ^{medium}
data points = 6

$$\text{entropy} = 0.918$$

Speaking	poor	avg	excellent	entropy
bad ^{hesitant}	2	0	0	0
good ^{fair}	0	4	0	0
v. good ^{fluent}	0	0	0	0

$$I = 0 \quad \text{gain} = 0.918$$

Speets	poor	avg	excellent	entropy
bad	2	2	0	1
good	0	1	0	0
v. good	0	1	0	0

$$I = \frac{4}{6} \times 1 = 0.666 \quad \text{gain} = 0.2514$$

Attribute	gain
Speaking	0.918
Speets	0.2514

There next node is 'speaking'!

Academy \rightarrow bad data points = 3

$$E(S) = 0.918$$

<u>Speaking</u>	poor	avg	excellent	entropy
bad	2	0	1	0.918
good	0	0	0	0
vgood	0	0	0	0

$$I = \frac{3}{3} \times 0.918 = .918$$

$$\text{gain} = E(S) - I = .918 - .918 = 0$$

<u>Creative</u>	poor	avg	excellent	entropy
bad	1	0	0	0
good	1	0	0	0
vgood	0	0	1	0

$$I = 0$$

$$\text{gain} = .918$$

<u>Sports</u>	poor	avg	excellent	entropy
bad	2	0	0	0
good	0	0	0	0
vgood	0	0	1	0

$$I = 0$$

$$\text{gain} = .918$$

Attri	gain
Speaking	0
Creative	.918
Sports	.918 ✓

on having a look at the data Sports is the apt next level node with highest gain