

Theory MT24007

Question - 1

Entropy of dataset: class | Approve: 5; Reject: 5

$$H(S) = -\frac{5}{10} \log_2\left(\frac{5}{10}\right) - \frac{5}{10} \log_2\left(\frac{5}{10}\right) = -0.5 \log_2(0.5) - 0.5 \log_2(0.5)$$

$$\boxed{H(S) = 1} \quad \checkmark$$

We have to find Information gain of each dataset attribute

Long-term Debt:

YES:

$$H(\text{YES}) = -\frac{1}{5} \log_2\left(\frac{1}{5}\right) - \frac{4}{5} \log_2\left(\frac{4}{5}\right)$$

$$\boxed{\text{YES} = 0.7219}$$

NO:

$$H(\text{NO}) = -\frac{4}{5} \log_2\left(\frac{4}{5}\right) - \frac{1}{5} \log_2\left(\frac{1}{5}\right)$$

$$\boxed{\text{NO} = 0.7219}$$

Combined entropy weightage:

$$H_{\text{Longterm}} = \frac{5}{10} \times 0.7219 + \frac{5}{10} \times 0.7219 = \underline{0.7219}$$

Information gain: $IG(\text{Longterm}) = 1 - 0.7219 = 0.2781$

$$\boxed{IG(\text{Longterm}) = 0.2781}$$

Unemployment:

$$H(\text{YES}) = -\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) = 0$$

$$H(\text{NO}) = -\frac{5}{8} \log_2\left(\frac{5}{8}\right) - \frac{3}{8} \log_2\left(\frac{3}{8}\right)$$

$$= 0.9544$$

weighted entropy: $H(\text{Unemployment}) = \frac{2}{10} \cdot 0 + \frac{8}{10} \cdot 0.9544 = \underline{0.7635}$

Information gain: $IG(\text{Unemployed}) = 1 - 0.7635 = \underline{0.2365}$

$$\boxed{IG(\text{Unemployed}) = 0.2365}$$

Credit Rating.

$$H(\text{Good}) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \\ = \underline{0.9183}$$

$$H(\text{BAD}) = -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) \\ = \underline{0.9852}$$

$$\text{Weighted: } H(\text{Credit Rating}) = \frac{3}{10} \cdot 0.9183 + \frac{7}{10} \cdot 0.9852 \\ = \underline{0.9645}$$

$$I G(\text{Credit Rating}) = 1 - 0.9645$$

$$\boxed{I G(\text{Credit Rating}) = 0.0355}$$

Down Payment < 20%.

$$H(\text{YES}) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \\ = -0.4 \log_2(0.4) - 0.6 (\log_2(0.6)) = \underline{0.97095}$$

$$H(\text{NO}) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \\ = 0.97095.$$

$$\text{Weighted: } H(\text{Down Payment} < 20\%) = \frac{5}{10} \cdot 0.97095 \times 2 \\ = \underline{0.97095}$$

$$I G : \quad \text{IG}(\text{Downpayment}) = 1 - 0.97095 \\ = 0.02905$$

$$\boxed{I G(\text{Downpayment} < 20\%) = 0.02905}$$

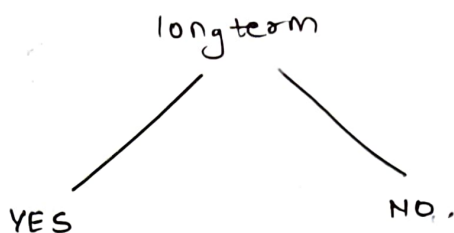
$$\text{long term: } 0.27813 \quad \underline{0.28}$$

$$\text{unemployed: } \underline{0.24}$$

$$\text{Credit Rating: } \underline{0.036}$$

$$\text{Down payment: } \underline{0.029}$$

longterm debt will be root node
[has highest gain 1G: 0.28]



Unemployed	Credit Rating	Down Payment	Class
No	Good	No	Approve
No	Bad	No	Reject
No	Bad	Yes	Reject
No	Bad	Yes	Reject
YES	Bad	No	Reject

longterm | YES

Unemployed: $E(\text{YES}) = -\frac{0}{1} \dots -\frac{1}{1} \log_2\left(\frac{1}{1}\right)$
 $= 0$

$E(\text{NO}) = -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right)$
 $= 0.244$

$1G = 1 - \left(\frac{1}{5} \cdot 0 + \frac{4}{5} \cdot 0.244\right) = 0.81$

$1G(\text{Unemployed}) = 0.81$

Credit Rating: $E(\text{Yes}) = E(\text{NO}) = 0$

$1G(\text{Credit R}) = 1$ \checkmark

Downpayment $E(\text{Yes}) = 0$

$E(\text{NO}) = -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right)$
 ≈ 0.477

$1G = 1 - \left(\frac{3}{5} \cdot 0.477\right) = 0.71$

$1G(\text{down payment}) = 0.71$

Unemployed	Credit Rating	Down Payment	Class
No	Good	Yes	Approve
No	Bad	No	Approve
No	Bad	Yes	Approve
No	Bad	No	Approve
YES	Good	Yes	Reject

longterm | NO

Unemployed: $E(\text{YES}) = 0$
 $E(\text{NO}) = 0$

$1G(\text{Unemployed}) = 1$ \checkmark highest

Credit Rating

$E(\text{Good}) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 0$

$E(\text{Bad}) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \approx 0.478$

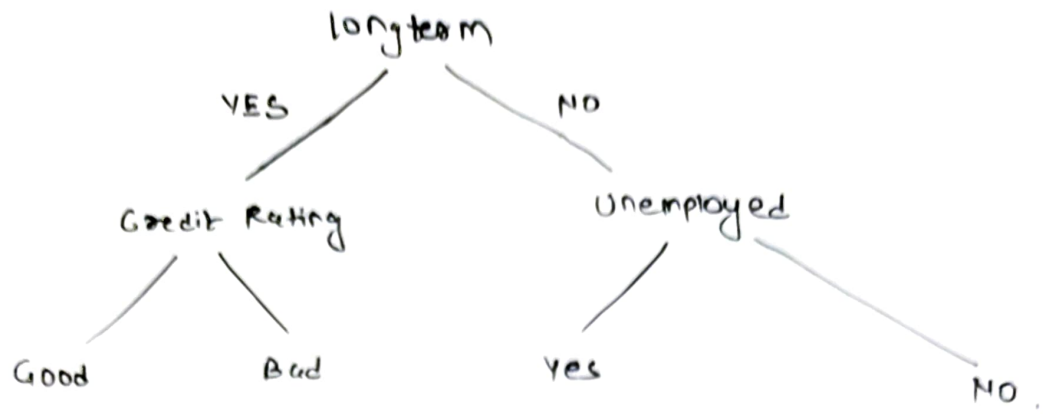
$1G(\text{Credit Rating}) = 0.71$

Downpayment

$E(\text{YES}) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.478$

$E(\text{NO}) = 0$

$1G(\text{Down Payment}) = 0.71$



Unemp	GR	DP	Class
NO	Good	NO	<u>Approve</u>
all			

Unemp	DP	Class
NO	NO	R
NO	YES	R
NO	YES	A
YES	NO	R
<u>all</u>		

G.R	D.P	Class
Good	YES	<u>Reject</u>
all		

G.R	D.P	Class
Good	YES	APP
Bad	NO	APP
Bad	YES	APP
Bad	NO	App
<u>all</u>		

FINAL DECISION TREE

