

Question - 1

Entropy of dataset : Class 1 Approve : 5 ; Reject : 5

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

H(S) = 1

We have to find Information gain of each dataset attribute

Long-term debt :

YES :

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

YES = 0.7219

NO :

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

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 : $IG(\text{longterm}) = 1 - 0.7219 = 0.2781$

IG (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 :

$$\underline{IG(\text{unemployed}) = 1 - 0.7635 = 0.2365}$$

IG (unemployed) = 0.2365

Credit Rating.

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

$= \underline{\underline{0.9183}}$

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

$= \underline{\underline{0.9852}}$

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

$= \underline{\underline{0.9645}}$

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

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

Down Payment < 20%.

$$H(YES) = -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \log_2 \left(\frac{3}{5}\right)$$

$= -0.4 \log_2(0.4) - 0.6 \log_2(0.6) = \underline{\underline{0.97095}}$

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

$= 0.97095$

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

$= \underline{\underline{0.97095}}$

$I_G : \quad I(G(Downpayment)) = 1 - 0.97095$

$= \underline{\underline{0.02905}}$

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

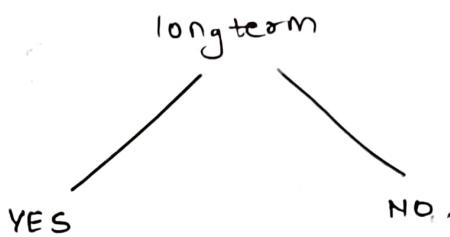
long term: $0.2781 \pm \underline{\underline{0.28}}$

Unemployed: $\underline{\underline{0.24}}$

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

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

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



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

longterm | YES

$$\# \underline{\text{Unemployed}} : E(\text{YES}) = -\frac{0}{1} - \frac{1}{1} \log\left(\frac{1}{1}\right) = 0$$

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

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

$$\boxed{IG(\text{Unemployed}) = 0.81}$$

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

$$\boxed{IG(\text{Credit R}) = 1} \Leftarrow$$

Downpayment : $E(\text{YES}) = 0$

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

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

$$\boxed{IG(\text{Down payment}) = 0.71}$$

longterm | NO

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

$$\boxed{IG(\text{Unemployed}) = 1} \Leftarrow \text{highest}$$

Credit Rating

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

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

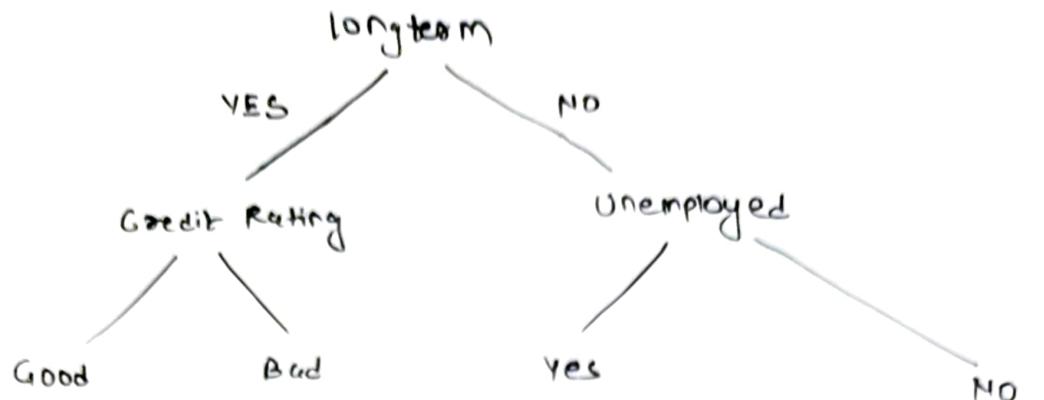
$$\boxed{IG(\text{Credit R}) = 0.71}$$

Down payment

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

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

$$\boxed{IG(\text{Down Payment}) = 0.71}$$



Unemp	GR	DP	Class	Unemp	DP	Class	GR	DP	Class	C.R.	D.P.	Class
No	Good	No	Approve	No	No	R	Good	Yes	Reject	Good	Yes	Appr.
				No	YES	R				Bad	No	Appr.
				No	YES	R				Bad	Yes	Appr.
				YES	NO	R				Bad	No	Appr.
												all

FINAL DECISION TREE

