Decision Tree: helps in taking decision

Dataset:

Age	Competition	Type	Profit
Old	Yes	s/w	Down
Old	No	s/w	Down
Old	No	h/w	Down
mid	Yes	s/w	Down
mid	Yes	h/w	Down
mid	No	h/w	Up
mid	No	s/w	Up
new	Yes	s/w	Up
new	No	h/w	Up
new	No	h/w	Up

Profit → Target Attribute.

(1) Calculate Information Gain (IG) of Target Att. i.e. Profit

$$IG = -\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

Note: P is Up $\therefore n(P) = 5$

N is Down $\therefore n(N) = 5$

$$\therefore IG = -\left[\frac{5}{10}\log_2\left(\frac{5}{10}\right) + \frac{5}{10}\log_2\left(\frac{5}{10}\right)\right]$$

$$= -(-1)$$

$$\boxed{IG = +1}$$

② Find Entropy of all Attributes

(a) Consider Age

	Down	Up
old	3	0
mid	2	2
new	0	3

Entropy = IG × Probability

old is Down 3 times

$$IG(old) = -\left[\frac{3}{3}\log_2\left(\frac{3}{3}\right) + \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right] = 0$$

total count of old in Age Column

$$IG(mid) = -\left[\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log\left(\frac{2}{4}\right)\right] = 1$$

$$I.G.(new) = -\left[\frac{0}{3}\log\left(\frac{0}{3}\right) + \frac{3}{3}\log\left(\frac{3}{3}\right)\right] = 0$$

$$\therefore E_{Age} = 0 \times \frac{3}{10} + 1 \times \frac{4}{10} + 0 \times \frac{3}{10}$$

$$\therefore \boxed{E(Age) = 0.4}$$

$$Gain_{(Age)} = IG_{Target} - E_{Age}$$

$$= 1 - 0.4$$

$$\boxed{Gain(Age) = 0.6}$$

Summarizing,

Gain(Age) = 0.60
Gain(Competition) = 0.124
Gain(Type) = 0
I.G.(Profit) = 1

Highest Gain = 0.6 ∴ Age is ROOT of Decision Tree.
2nd is Competition.