(5)

Entropy of target variable,

$$E = -P(yes)\log_2 P(yes) - P(No)\log_2 P(No) \quad -(i)$$

total No. of "yes" outcomes = 9

total No. of "No" outcomes = 5

$$E = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$$

Gain of outlook

total "Sunny" outlook = 5

No. of having "yes" = 2

No. of having "No" = 3

$$E_{Sunny} = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$$

total "overcast" outlook = 4

No. of "yes" = 4

No. of "No" = 0

$$E_{overcast} = -\frac{4}{4}\log\frac{4}{4} - \frac{0}{4}\log\frac{0}{4} = 0$$

total No. of "Rainy" outlook = 5

having "yes" = 3

having "No" = 2

$$E_{Rainy} = -\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5} = 0.971$$

$$Gain = 0.94 - \frac{5}{14}(0.971 + 0.971) = 0.94 - 1.942 \times \frac{5}{14}$$

$$Gain_{outlook} = 0.246$$

Gain of temp

$$E_{Hot} = -\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4} = 1$$

$$E_{mild} = -\frac{4}{6}\log\frac{4}{6} - \frac{2}{6}\log\frac{2}{6} = 0.9183$$

$$E_{cold} = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.8113$$

$$Gain_{temp} = 0.94 - \frac{4}{14}(1) - \frac{6}{14}(0.9183) - \frac{4}{14}(0.8113)$$

$$= 0.0289$$

Gain of Humidity

$$E_{High} = -\frac{3}{7}\log\frac{3}{7} - \frac{4}{7}\log\frac{4}{7} = 0.9852$$

$$E_{Normal} = -\frac{6}{7}\log\frac{6}{7} - \frac{1}{7}\log\frac{1}{7} = 0.5916$$

$$Gain_{Humidity} = 0.94 - \frac{7}{14}(0.9852) - \frac{7}{14}(0.5916)$$

$$= 0.1516$$

Gain of windy

$$E_{weak} = -\frac{6}{8}\log\frac{6}{8} - \frac{2}{8}\log\frac{2}{8} = 0.8113$$

$$E_{strong} = \frac{3}{6}\log\frac{3}{6} - \frac{3}{6}\log\frac{3}{6} = 1.0$$

$$Gain_{windy} = 0.94 - \frac{6}{14}(0.8113) - \frac{8}{14}(1)$$

$$= 0.0478$$

$$Gain_{outlook} > Gain_{Humidity} > Gain_{windy} > Gain_{temp}$$

∴ Root Node = outlook

① Overcast

From data given, whenever outlook is overcast, there is positive outcomes.

∴ overcast mean match will happen.

② Sunny

$$E_{sunny} = 0.971$$

$$Gain_{temp} = 0.971 - \left(E_{Hot} \times \frac{2}{5} + \frac{2}{5} E_{mild} + \frac{1}{5} E_{cool} \right)$$

$$Gain_{temp} = 0.971 - \left(\frac{2}{5}\times(0) + \frac{2}{5}\times(1) + \frac{1}{5}\times 0\right)$$

$$= 0.971 - 0.4 = 0.571$$

$$Gain_{Humidity} = 0.971 - \left(\frac{3}{5} E_{high} + \frac{2}{5} E_{Normal}\right)$$

$$= 0.971 - \left(\frac{3}{5}\times(0) + \frac{2}{5}\times(0)\right) = 0.971$$

$$Gain_{wind} = 0.971 - \left(\frac{3}{5} E_{weak} + \frac{2}{5} E_{strong}\right)$$

$$= 0.971 - \left(\frac{3}{5}\times\left(-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right) + \frac{2}{5}\times(1)\right).$$

$$= 0.0192$$

Gain of Humidity maximum.

⇒ So whenever outlook is sunny, Humidity is High, match will not happen.
And Humidity Normal match will happen.

(3) Rainy

$$E_{rainy} = 0.971$$

$$Gain_{temp} = 0.971 - \left(\frac{3}{5} E_{mild} + \frac{2}{5} E_{cool}\right)$$

$$= 0.971 - \left(\frac{3}{5}\times\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right) + \frac{2}{5}\times(1)\right)$$

$$= 0.0192$$

$$Gain_{Humidity} = 0.971 - \left(\frac{2}{5} E_{High} + \frac{3}{5} E_{Normal}\right)$$

$$= 0.971 - \left(\frac{2}{5} \times (1) + \frac{3}{5} \times \left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right)\right)$$

$$= 0.019_2$$

$$Gain_{wind} = 0.971 - \left(\frac{3}{5} \times E_{weak} + \frac{2}{5} E_{strong}\right)$$

$$= 0.971 - \left(\frac{3}{5} \times (0) + \frac{2}{5} \times (0)\right)$$

$$= 0.971$$

$Gain_{wind}$ maximum.

when outlook is Rainy —

(1) match will happen if, wind is weak

(2) match will not happen, if wind is strong

overall representation