Entropy Algorithm - $\sum_{i=1}^{n} P_i \log_2 P_i$

1. Probability of the "will Excercise" :-

P(Yes) = 4 , P(No) = 3, Total = 7

Entropy = $-(4/7) \times \log_2(4/7) - (3/7) \times \log_2(3/7)$

$= -0.46 - 0.52$

$= 0.98$

2. Entropy for value of weather (Sunny & Rainy)

Sunny - P(Yes) = 3 , P(No) = 0 ,

Entropy of Sunny = 0 , Since there is no subset of data.

Rainy - P(Yes) = 1 , P(No) = 3.

Entropy of Rainy = $-(\frac{1}{4}) \times \log_2(\frac{1}{4}) - (\frac{3}{4}) \times \log_2(\frac{3}{4})$

$= -0.5 - 0.31$

Rainy = 0.81

3. Information Gain of weather : TE - WE

weighted Entropy = $\frac{3}{7} \times$ (Sunny) + $\frac{4}{7}$ (Rainy)

$= \frac{3}{7} \times 0 + \frac{4}{7} \times 0.81$

WE = 0.522

Weather = 0.98 − 0.522

= 0.458.

=

4. Information Gain for Just Ate:

$P(Yes) = -\frac{2}{3}\left[\log\left(\frac{2}{3}\right)\right] - \frac{1}{3}\left[\log\left(\frac{1}{3}\right)\right]$

$= \frac{1.16}{3} + \frac{1.58}{3} = 0.91$

$P(No) = -\frac{2}{4}\left[\log\left(\frac{2}{4}\right)\right] - \frac{2}{4}\left[\log\left(\frac{2}{4}\right)\right]$

$= \frac{1}{2} + \frac{1}{2} = 1$

Gain of Just Ate = TE − JAE

$0.98 - \left[\frac{3}{7}(0.91) + \frac{4}{7}(1)\right]$

$0.98 - 0.96$

Just Ate = 0.02

=

5. Information Gain for Work Late:

$P(Yes) = -\frac{1}{3}\left[\log\left(\frac{1}{3}\right)\right] - \left[\frac{2}{3}\log\left(\frac{2}{3}\right)\right]$

$= -0.53 - 0.38$

$= 0.91$

$P(No) = -\frac{3}{4}\left[\log\left(\frac{3}{4}\right)\right] - \frac{1}{4}\left[\log\left(\frac{1}{4}\right)\right]$

$= -0.31 - 0.5$

$= 0.81$

Gain for Work Late = TE - WLE

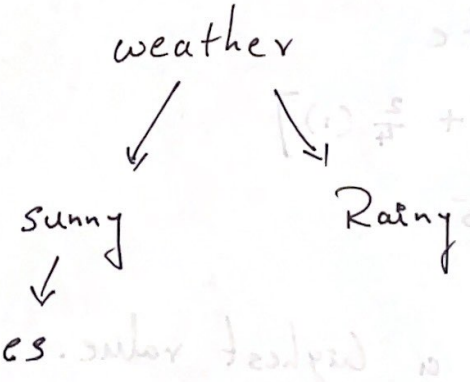
$$0.98 - \left[\frac{3}{7}(0.91) + \frac{4}{7}(0.81)\right]$$

$$0.98 - 0.39 + 0.46$$

$$0.98 - 0.85$$

Work Late = 0.13.

So, Weather has a highest Entropy value, Taking weathe as Root Node



So, Now we have left Just Ate & Work Late Attributes

$$P(\text{Rainy weather}) = -\frac{3}{4}\left(\log\left(\frac{3}{4}\right)\right) - \frac{1}{4}\left(\log\left(\frac{1}{4}\right)\right)$$

$$= 0.81$$

Entropy of Just Ate

$$P(\text{Yes}) = 0$$

$$P(\text{No}) = -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) = 0.91$$

Gain of Just Ate

$$0.81 - \left[\frac{1}{4}(0) + \frac{3}{4}(0.91)\right]$$

$$= 0.13$$

Entropy of Work Late

$$P(\text{Yes}) = 0$$

$$P(\text{No}) = -\frac{1}{2}\left[\log\left(\frac{1}{2}\right)\right] - \frac{1}{2}\left[\log\left(\frac{1}{2}\right)\right]$$

$$= 1$$

Gain of Work Late

$$0.81 - \left[0 + \frac{2}{4}(1)\right]$$

$$= 0.80 - 0.5$$

$$= 0.31$$

Now work late has a highest value.

Decission Tree:-