# Decision Tree

Data

| weather | Humidity | wind | play |
|---|---|---|---|
| Sunny | High | weak | No |
| Sunny | Normal | weak | No |
| Sunny | Normal | Strong | No |
| Cloudy | High | weak | NO |
| Cloudy | High | Strong | NO |
| Cloudy | Normal | Strong | yes |
| Cloudy | Normal | weak | yes |
| Rainy | High | weak | yes |
| Rainy | Normal | strong | yes |
| Rainy | Normal | weak | yes |

Steps

i) Choosing Target Attribute (T.A)

ii) Calculating Information Gain (I.G) of T.A.

$$I.G = \frac{-P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

iii) Entropy of other Attributes, with

$$E(A) = \sum_{i=1}^{N} \frac{P_i + N_i}{P+N} \times I(P_i N_i)$$

iv) substracting E(A) from I.G of each Attribute for Gain for particular attributes. (Root, leaf, child etc [illegible])

## Formulas

$$I.G = \frac{-P}{P+N} \log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N} \log_2\left(\frac{N}{P+N}\right)$$

$$E(A) = \sum_{i=1}^{N} \frac{Pi + Ni}{P+N} \times I(Pi, Ni)$$ (probability)

$$Gain = I.G - E(A)$$

## Solve:

Lets assume Target Attribute is play.

∴ I.G for Play:

$$I.G = -\frac{P}{P+N} \log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N} \log_2\left(\frac{N}{P+N}\right)$$

$$= -\frac{5}{5+5} \log_2\left(\frac{5}{5+5}\right) - \frac{5}{5+5} \log_2\left(\frac{5}{5+5}\right)$$

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

$$= -1/2 \log_2(2^{-1}) - 1/2 \log_2(2^{-1})$$

$$= 1/2 + 1/2$$

$$= 1$$

Weather

| | No | Yes |
|---|---|---|
| Sunny | 3 | 0 |
| Cloudy | 2 | 2 |
| Rainy | 0 | 3 |

$E(A).\ Sunny = -\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{N+P}\log_2\left(\frac{N}{P+N}\right) \times I.P;N$

$= -3/3\ \log_2(3/3) - 0/3\ \log_2(0/3) \times 3/10$

$= 0 \times 3/10$

$= 0$

$E(A).\ Cloudy = -2/4\ \log_2(2/4) - 2/4\ \log_2(2/4) \times 4/10$

$= -1/2\ \log_2(1/2) - 1/2\ \log(1/2) \times \frac{4}{10}$

$= 1 \times \frac{4}{10} = 0.4$

$E(A).\ Rainy = -0/3\ \log_2(0/3) - 3/3\ \log_2(3/3) \times 3/10$

$= 0 \times 3/10$

$= 0$

$\therefore E(\text{weather}) = 0 + 0.4 + 0 = 0.4$

$\therefore Gain(\text{weather}) = 1 - 0.4 = 0.6$

## Humidity

| | No | Yes |
|---|---|---|
| High | 3 | 1 |
| Normal | 2 | 4 |

$E(\text{High}) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) \times \frac{4}{10}$

$= 0.8113 \times \frac{4}{10} = 0.325$

$E(\text{Normal}) = \frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{4}{6}\log_2\left(\frac{4}{6}\right) \times \frac{6}{10}$

$= 0.9183 \times \frac{6}{10} = 0.551$

$\therefore E(\text{Humidity}) = 0.325 + 0.551 = 0.876$

$\therefore \text{Gain}(\text{Humidity}) = 1 - 0.876 = 0.124$

## Wind

| | No | Yes |
|---|---|---|
| weak | 3 | 3 |
| Strong | 2 | 2 |

$E(\text{weak}) = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right) \times \frac{6}{10}$

$= 1 \times \frac{6}{10} = 0.6$

$E(\text{Strong}) = \frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) \times \frac{4}{10}$

$= 1 \times \frac{4}{10} = 0.4$

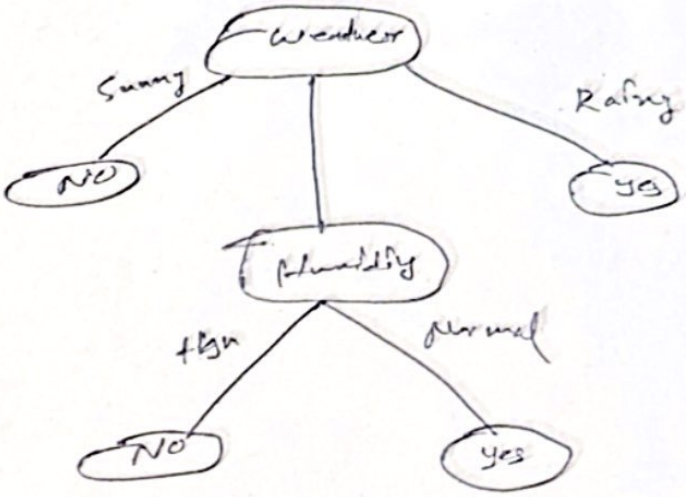
$\therefore E(\text{wind}) = 0.6 + 0.4 = 1$

$\therefore \text{Gain}(\text{wind}) = 1 - 1 = 0$

∴ Decision Tree

Here,

Gain (Weather) = 0.6

Gain (Humidity) = 0.124

Gain (Wind) = 0