Decision Tree Assignment
(submitted by -Abhishek Prabhudesai)
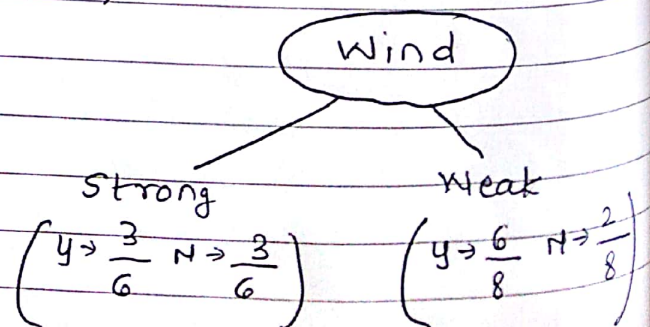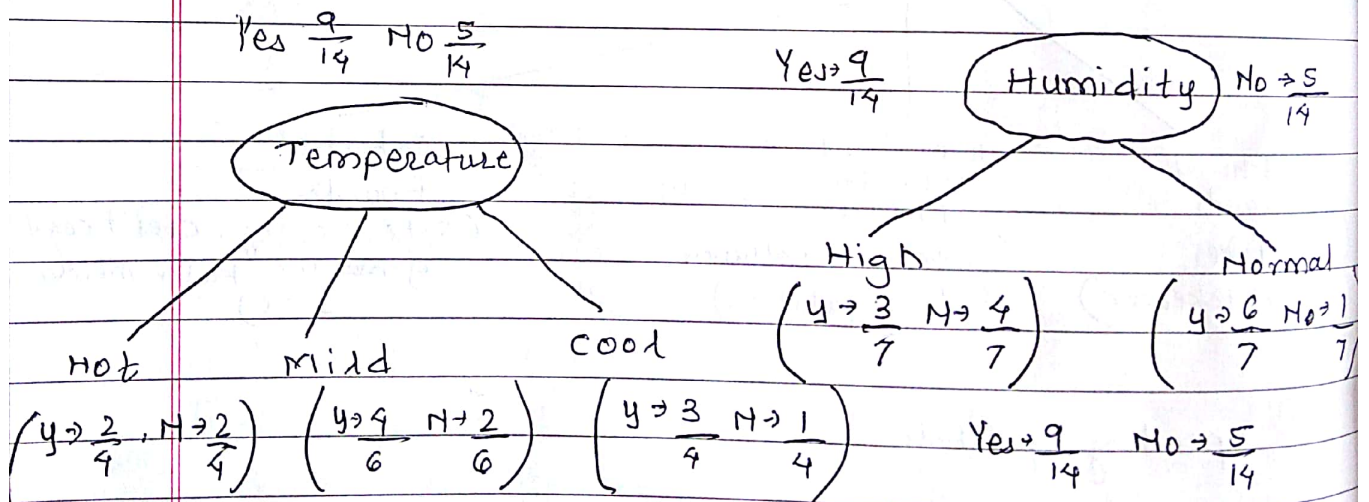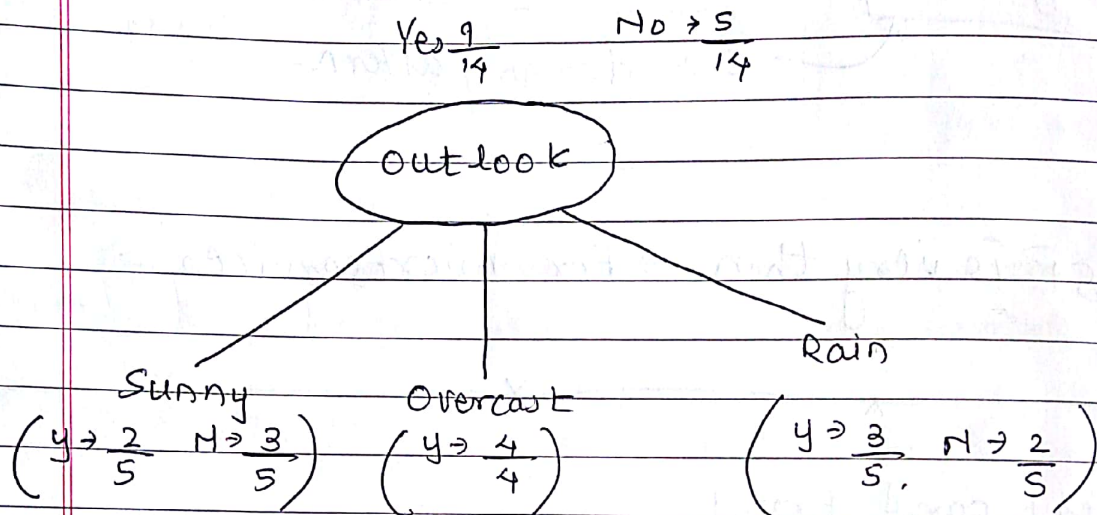
Please develop decision tree for ICC World Cup 2015 Play Cricket.

— Decision Tree Calculations

There are 4 different columns (attributes)

① Outlook — Has 3 tuples — Sunny, Overcast, Rain

② Temperature — Has 3 tuples — Hot, Mild, Cool

③ Humidity — Has 2 tuples — High, Normal

④ Wind — Has 2 tuples — Strong, Weak

Yes $\frac{9}{14}$ No $\to \frac{5}{14}$

Outlook

Sunny $\left(y \to \frac{2}{5} \quad N \to \frac{3}{5} \right)$

Overcast $\left(y \to \frac{4}{4} \right)$

Rain $\left(y \to \frac{3}{5}, \quad N \to \frac{2}{5} \right)$

Yes $\frac{9}{14}$ No $\frac{5}{14}$

Temperature

Hot $\left(y \to \frac{2}{4}, N \to \frac{2}{4} \right)$

Mild $\left(y \to \frac{4}{6} \quad N \to \frac{2}{6} \right)$

Cool $\left(y \to \frac{3}{4} \quad N \to \frac{1}{4} \right)$

Yes $\to \frac{9}{14}$ **Humidity**   No $\to \frac{5}{14}$

High $\left(y \to \frac{3}{7} \quad N \to \frac{4}{7} \right)$

Normal $\left(y \to \frac{6}{7} \quad N \to \frac{1}{7} \right)$

Yes $\to \frac{9}{14}$ No $\to \frac{5}{14}$

Wind

Strong $\left(y \to \frac{3}{6} \quad N \to \frac{3}{6} \right)$

Weak $\left(y \to \frac{6}{8} \quad N \to \frac{2}{8} \right)$

● step (1) → Calculate Entropy of entire dataset

$$\text{Entropy } (D) = -\sum_{i=1}^{m} P_i \log_2 (P_i)$$

$$= -\frac{9}{14} \log_2 \left(\frac{9}{14}\right) - \frac{5}{14} \log_2 \left(\frac{5}{14}\right)$$

$$= 0.940286 \text{ bit.}$$

● step (2) - Calculate entropy of tuples

Outlook

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

$$= 0.970950 \text{ bit}$$

$$H\left(S_{overcast}\right) = \frac{-4}{4} \log_2 \left(\frac{4}{4}\right)$$

$$= 0 \text{ bit}$$

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

$$= 0.970950 \text{ bit}$$

Temperature

$$H\left(S_{hot}\right) = \frac{-2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right)$$

$$= 1 \text{ bit}$$

$$H\left(S_{mild}\right) = \frac{-4}{6} \log_2 \left(\frac{4}{6}\right) - \frac{2}{6} \log_2 \left(\frac{2}{6}\right)$$

$$= 0.918295 \text{ bit.}$$

$$H(S_{cool}) = \frac{-3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right)$$

$$= 0.811 \text{ bit}$$

Wind

$$H(S_{strong}) = \frac{-3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right)$$

$$= 1 \text{ bit}$$

$$H(S_{weak}) = \frac{-6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right)$$

$$= 0.811278 \text{ bit}$$

Humidity

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

$$= 0.985228 \text{ bit}$$

$$H(S_{Normal}) = \frac{-6}{7} \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \log_2\left(\frac{1}{7}\right)$$

$$= 0.591672 \text{ bit}$$

Step (3) - Calculate gain

$$\text{Gain}(S, A) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)$$

① Gain (S, outlook) = $0.9403 - \dfrac{5}{14} \times 0.9710 * - \dfrac{4}{14} \times 0$

$$- \dfrac{5}{14} \times 0.9710$$

$$= 0.2467 \text{ bit}$$

② Gain (s, temp) = $0.9403 - \dfrac{4}{14} \times 1 - \dfrac{6}{14} \times 0.9183$

$$- \dfrac{4}{14} \times 0.811$$

$$= 0.0293 \text{ bit}$$

③ Gain (S, wind) = $0.9403 - \dfrac{6}{14} \times 1 - \dfrac{8}{14} \times 0.8113$

$$= 0.0481 \text{ bit}$$

④ Gain (S, Humidity) = $0.9403 - \dfrac{7}{14} \times 0.9852$

$$- \dfrac{7}{14} \times 0.5917$$

$$= 0.1519 \text{ bit}$$

• Step (4) - Arrange gain by desc. order.

Outlook
Humidity
Wind
temp.

- Step (s) – Develop decision tree

```
                    ( Outlook )
          sunny    /     |        \  Rain
                  /       | Overcas  \
            [Humidity]   [Yes]      [Wind]
    High   /      \ Normal    Weak  /      \ Strong
      [No]    [Yes]            [Yes]      [No]
```