Assignment - 2

Q-3 Let us - - - - - - - - - . decision tree.

Soln Play cricket → target column.

Attributes out of which we select the root node.

1) Outlook
2) Temperature
3) Humidity
4) Wind

Info gain = Entropy (Before) - Entropy (After)

E(S) : Entropy of dataset.

$PC_+ = 9$

$PC_- = 5$

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

$$\boxed{E(S) = 0.940}$$

E (O) = ? → Entropy of Outlook.

Outlook → Sunny
→ Overcast
→ Rain

PAGE NO.:
DATE: / /

$$E(O) = E(S) + E(O) + E(R)$$

E(O) → entropy of outlook; E(S) → entropy of sunny; E(O) → entropy of overcast; E(R) → entropy of rain ---- (*) eq.

so, $E(S) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$

$E(S) = 0.97$ ---- (1) eq.

$E(O) = -\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right)$

$E(O) = 0$ ---- (2) eq

$E(R) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$

$E(R) \Rightarrow 0.97$ ---- (3) eq

~~adding~~ putting eq (1), (2) & (3) in (*),

$E(O) = 1.94$

E(O) → Entropy of outlook

Information gain of outlook $G(O) = E(S) - \sum_{v \in V} \frac{|S_v|}{S} \varepsilon(S_v)$

$$= 0.94 - \underbrace{\frac{5}{14} \times 0.971}_{\text{Sunny}} - \underbrace{\frac{4}{14} \times (0)}_{\text{Overcast}} - \underbrace{\frac{5}{14} \times (0.971)}_{\text{Rain}}$$

$G(O) = \underline{\underline{0.207}}$

Hot entropy Mild entropy

$E(\text{Temperature})$ or $E(T) = E(H) + E(M) + E(C)$

entropy of cool.

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

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

$$E(M) = -\frac{4}{6} \log_2 \left(\frac{4}{6}\right) - \frac{2}{6} \log_2 \left(\frac{2}{6}\right)$$

$$= 0.918$$

$G(W) = 0.048$

$G(\text{Outlook}), G(\text{Temp}), G(\text{Hum}), G(W)$

As $G(\text{Outlook})$ is maximum i.e

Outlook will be the root node.

Q-4 ~~[illegible]~~ K-NN.

$x_1 = 161, \quad y_1 = 61.$

$D_1 = \sqrt{(161-158)^2 + (61-58)^2} = 4.25$

$D_2 = \sqrt{(161-158)^2 + (61-59)^2} = 3.6$

$D_3 = \sqrt{(161-158)^2 + (61-63)^2} = 3.6$

$D_4 = \sqrt{(161-160)^2 + (61-59)^2} = 2.2$

$D_5 = \sqrt{(161-160)^2 + (161-60)^2} = 1.4$

$D_6 = \sqrt{(161-163)^2 + (161-60)^2} = 2.2$

$D_7 = \sqrt{(161-163)^2 + (61-61)^2} = 2$

$D_8 = \sqrt{(161-160)^2 + (61-64)^2} = 3.2$

$D_9 = \sqrt{(161-163)^2 + (61-64)^2} = 3.6$

$D_{10} = \sqrt{(161-165)^2 + (61-61)^2} = 4.0$

$D_{11} = \sqrt{(161-165)^2 + (61-62)^2} = 4.1$

$D_{12} = \sqrt{(161-165)^2 + (61-65)^2} = 5.7$

$D_{13} = \sqrt{(161-168)^2 + (61-62)^2} = 7.1$

$D_{14} = \sqrt{(161-168)^2 + (61-63)^2} = 8.6$

219

$D_{15} = \sqrt{(161-168)^2 + (61-66)^2} = 9.4$

219

$D_{16} = \sqrt{(161-170)^2 + (61-63)^2} = 9.$

$D_{17} = \sqrt{(161-170)^2 + (61-64)^2} = 9.486$

$D_{18} = \sqrt{(161-170)^2 + (61-68)^2} = 11.401$

PAGE NO.:
DATE: / /

Let $k=3$

D_4, D_5 & D_7.

Hence, your neighbour contains T-shirt size = M, M, M.

∴ Our model will classify Monica's T-shirt size as M

Q-5