

## Enyia Esther - Workshop 7

## Theoretical Task T7.1 Solution

Given Data:

$X_1$	$X_2$	$Y$
1	1	1
1	1	1
1	1	2
1	0	3
0	0	2
0	0	3

**Step 1: Calculate Initial Entropy  $H(Y)$** 

Class frequencies:

- $P(Y = 1) = \frac{2}{6}$
- $P(Y = 2) = \frac{2}{6}$
- $P(Y = 3) = \frac{2}{6}$

$$H(Y) = - \sum_{i=1}^3 P(Y = i) \log_2 P(Y = i) = - \left( \frac{2}{6} \log_2 \frac{2}{6} + \frac{2}{6} \log_2 \frac{2}{6} + \frac{2}{6} \log_2 \frac{2}{6} \right) = \log_2 3 \approx 1.585 \text{ bits}$$

**Step 2: Information Gain for  $X_1$** Split on  $X_1$ :

1.  $X_1 = 1$  **branch (4 instances):**
  - $Y = 1: 2, Y = 2: 1, Y = 3: 1$

- $H(Y|X_1 = 1) = -\left(\frac{2}{4}\log_2 \frac{2}{4} + \frac{1}{4}\log_2 \frac{1}{4} + \frac{1}{4}\log_2 \frac{1}{4}\right) = 1.5 \text{ bits}$

2.  $X_1 = 0$  **branch (2 instances):**

- $Y = 2: 1, Y = 3: 1$
- $H(Y|X_1 = 0) = -\left(\frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{2}\log_2 \frac{1}{2}\right) = 1 \text{ bit}$

**Weighted average entropy:**

$$H(Y|X_1) = \frac{4}{6} \times 1.5 + \frac{2}{6} \times 1 = 1.\overline{3} \text{ bits}$$

**Information Gain:**

$$IG(X_1) = H(Y) - H(Y|X_1) = 1.585 - 1.333 = 0.252 \text{ bits}$$

### Step 3: Information Gain for $X_2$

**Split on  $X_2$ :**

1.  $X_2 = 1$  **branch (3 instances):**

- $Y = 1: 2, Y = 2: 1$
- $H(Y|X_2 = 1) = -\left(\frac{2}{3}\log_2 \frac{2}{3} + \frac{1}{3}\log_2 \frac{1}{3}\right) \approx 0.918 \text{ bits}$

2.  $X_2 = 0$  **branch (3 instances):**

- $Y = 2: 1, Y = 3: 2$
- $H(Y|X_2 = 0) = -\left(\frac{1}{3}\log_2 \frac{1}{3} + \frac{2}{3}\log_2 \frac{2}{3}\right) \approx 0.918 \text{ bits}$

**Weighted average entropy:**

$$H(Y|X_2) = \frac{3}{6} \times 0.918 + \frac{3}{6} \times 0.918 = 0.918 \text{ bits}$$

**Information Gain:**

$$IG(X_2) = H(Y) - H(Y|X_2) = 1.585 - 0.918 = 0.667 \text{ bits}$$

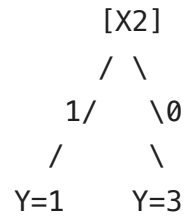
### Step 4: Comparison

- $IG(X_1) = 0.252 \text{ bits}$

- $IG(X_2) = 0.667$  bits

$X_2$  provides higher information gain → Better split

## Step 5: Decision Tree Structure



### Leaf predictions:

- $X_2 = 1$  **branch**: Predict  $Y = 1$  (majority class: 2/3 instances)
- $X_2 = 0$  **branch**: Predict  $Y = 3$  (majority class: 2/3 instances)

