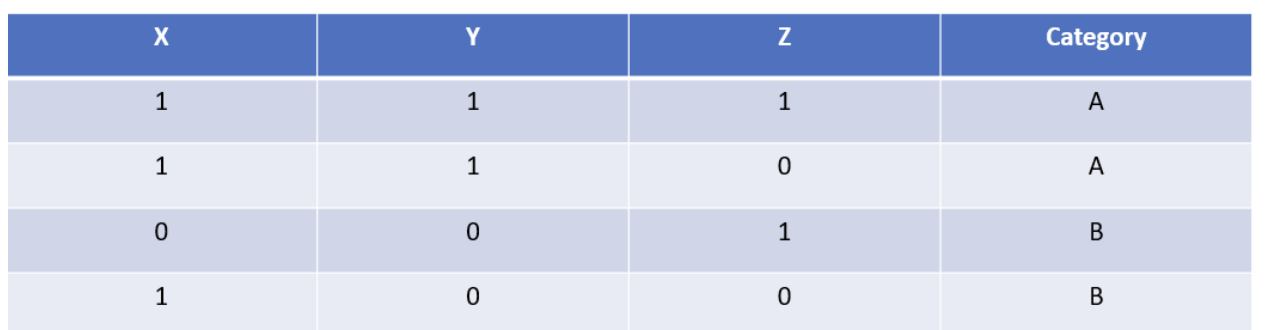
### **Entropy and Information Gain to determine which of X, Y and Z would be the best choice of Root node for the Decision Tree algorithm**

We have the following dataset:



**1. The Overall Entropy of the Dataset:**

**Total records**: N = 4

* **Class distribution**:  
  + **P:** 2 occurrences
  + **Q:** 2 occurrences

H(D)=−p log2(p) - q log2 (q)

=−2/4​×log2​(2/4​) - 2/4​×log2​(2/4​)

=−(0.5×log2​(0.5)+0.5×log2​(0.5))

=1.0

**Overall Entropy: H(D)=1.0**

### **2. Information Gain for Each Attribute:**

### **Information Gain for X:**

* **Splitting by X:**
  + X=1, 3 records → [A, A, B]
  + X=0, 1 record → [B]

#### **Entropy of Subsets:**

* For X=1

H(X=1) = −p log2(p) - q log2 (q)

= -(2/3×log⁡2(2/3) + 1/3×log⁡2(1/3))

=−(0.6667×−0.584+0.3333×−1.585)

=0.9183

* For X=0

H(X=0)= −p log2(p) - q log2 (q)

=−(1×log⁡2(1))

=0

H(X = 0) = 0

#### **Weighted Entropy:**

H(X)=3/4×0.9183+1/4×0

H(X) = 0.6887

#### **Information Gain: (overall entropy - weighted entropy)**

IG(X)=H(D)−H(X)

IG(X) = H(D) - H(X) =1.0−0.6887=0.3113

**Information Gain for Y:**

* **Splitting by Y:**
  + Y=1, 2 records → [A, A]
  + Y=0, 2 records → [B, B]

#### **Entropy of Subsets:**

* For Y=1

H(Y=1)=−(1×log⁡2(1))

=0

H(Y = 1) = 0

* For Y=0

H(Y=0)=−(1×log⁡2(1))

=0

H(Y = 0) = 0

#### **Weighted Entropy:**

H(Y)=2/4×0+2/4×0

=0

H(Y) = 0

#### **Information Gain:**

IG(Y) = H(D)−H(Y)

IG(Y) = 1.0−0

=1.0

**Information Gain for Z:**

* **Splitting by Z:**
  + Z=1, 2 records → [A, B]
  + Z=0, 2 records → [A, B]

#### **Entropy of Subsets:**

* For Z=1

H(Z=1)=−(1/2×log⁡2(1/2) + 1/2×log⁡2(1/ 2))

H(Z = 1) = 1.0

* For Z=0

H(Z = 0) = −(1/2×log⁡2(1/2) + 1/2×log⁡2(1/ 2))

H(Z = 0) = 1.0

#### **Weighted Entropy:**

H(Z)=2/4×1.0+2/4×1.0

H(Z) = 1.0

#### **Information Gain:**

IG(Z) = H(D)−H(Z)

IG(Z) = H(D) - H(Z) =1.0−1.0 = 0

### **Final Results:**

* **Overall Entropy:** 1.0
* **Information Gain for X:** 0.3113
* **Information Gain for Y:** 1.0
* **Information Gain for Z:** 0.0

### **Conclusion:**

Since **Y** has the **highest Information Gain (1.0)**, it is the **best choice for the root node** in the decision tree.