**Entropy:**

Entropy, as it relates to machine learning, is a measure of the randomness in the information being processed. The higher the entropy, the harder it is to draw any conclusions from that information.

Information gain (I (**p,n)) : [ (-p/p+n) \* ( log(2) p/p+n) - (n/p+n) \* (log(2) n/p+n) ]**

**Entropy** = ∑ P[i]+N[i] (I(P[i],N[i])

P+N

**Gain:** Entropy (class) - Entropy (attribute)

**Ex:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **Competition** | **Type** | **Profit** |
| Old | Yes | software | Down |
| Old | No | Software | Down |
| Old | No | Hardware | Down |
| Mid | Yes | Software | Down |
| Mid | Yes | Hardware | Down |
| Mid | No | Hardware | Up |
| Mid | No | Software | Up |
| New | Yes | Software | Up |
| New | No | Hardware | Up |
| New | No | Software | Up |

**Profit (class):** P = 5 (Up), N = 5 (Down)

**Class Entropy**:

-5/10 log2 (5/10) - 5/10 log2 (5/10) = 1

**Steps to calculate making decision tree:**

1. Information Gain
2. Entropy
3. Gain

**Age:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **P[i]** | **N[i]** | **I(P[i],N[i])** |
| Old | 0 | 3 | 0 |
| Mid | 2 | 2 | 1 |
| New | 3 | 0 | 0 |

**Entropy (Age):** 0+3/10 \* (0) + 2+2/10 \* (1) + 3+0/10 \* (0)

4/10=0.4

**Gain:**  1-0.4 = 0.6

**Competition:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Competition** | **P[i]** | **N[i]** | **I(P[i],N[i])** |
| Yes | 1 | 3 | 0.8127 |
| No | 4 | 2 | 0.918295 |

**Entropy (Competition):** 1+3/10 \* (0.8127) + 4+2/10 \* (0.918295) = 0.8754

**Gain:** 1- 0.8754 = 0.124515

**Type:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **P[i]** | **N[i]** | **I(P[i],N[i])** |
| Software | 3 | 3 | 1 |
| Hardware | 2 | 2 | 1 |

**Entropy (Type):** 3+3/10 \* (1) + 2+2/10 \* (1) = 1

**Gain:** 1-1 = 0

|  |  |
| --- | --- |
| **Attribute** | **Gain** |
| Age | 0.6 |
| Competition | 0.124 |
| Type | 0 |

**Compare Gain:**

Here Age gain is more, so Age is the Root node

Old

New

Mid

Down ? Up

**Age:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **Competition** | **Type** | **Profit** |
| Mid | Yes | Software | Down |
| Mid | Yes | Hardware | Down |
| Mid | No | Software | Up |
| Mid | No | Hardware | Up |

**Profit (class):** p = 2, n = 2

**Entropy:**

-2/4 log2 (2/4) - 2/4 log2 (2/4) = 1

**Competition:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Competition** | **P[i]** | **N[i]** | **I(P[i],N[i])** |
| Yes | 0 | 2 | 0 |
| No | 2 | 0 | 0 |

**Entropy (Competition) :**

0+2/4 (0) + 2+0/4 (0) = 0

**Gain:** Entropy (class) - Entropy (attribute)

1-0 =1

**Type:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **P[i]** | **N[i]** | **I(P[i],N[i])** |
| Software | 1 | 1 | 1 |
| Hardware | 1 | 1 | 1 |

**Entropy (Type) :**1+1/4 (1) + 1+1/4 (1) = 1

**Gain:** Entropy (class) - Entropy (attribute) = 1-1 =0

**Compare Gain:**

|  |  |
| --- | --- |
| **Attribute** | **Gain** |
| Competition | 1 node |
| Type | 0 |

**Decision Tree:**

Old

new

mid

Down Up

Competition

Yes No

**Down** **Up**