Manual Calculation of Decision Tree using Information Gain (IG) and Gini Index (GI)

# Step 1: Recall the formulas

Entropy & Information Gain:

Entropy(S) = - Σ p\_i log2 p\_i

IG(S, A) = Entropy(S) - Σ (|S\_v|/|S| \* Entropy(S\_v))

Gini Index:

Gini(S) = 1 - Σ p\_i^2

Gini\_index(A) = Σ (|S\_v|/|S| \* Gini(S\_v))

# Step 2: Example Subset of German Credit Dataset

|  |  |  |
| --- | --- | --- |
| Applicant | checking\_status | Class |
| 1 | <0 DM | Bad |
| 2 | <0 DM | Bad |
| 3 | <0 DM | Good |
| 4 | 0–200 DM | Good |
| 5 | 0–200 DM | Good |
| 6 | 0–200 DM | Bad |
| 7 | no checking | Good |
| 8 | no checking | Good |
| 9 | no checking | Bad |
| 10 | no checking | Good |

Totals: 6 Good, 4 Bad

# Step 3: Entropy of the whole dataset

p\_good = 6/10 = 0.6, p\_bad = 4/10 = 0.4

Entropy(S) = -(0.6 log2 0.6 + 0.4 log2 0.4) = 0.971

# Step 4: Entropy of subsets for checking\_status

<0 DM: (1 Good, 2 Bad) → Entropy = 0.918

0–200 DM: (2 Good, 1 Bad) → Entropy = 0.918

no checking: (3 Good, 1 Bad) → Entropy = 0.811

# Step 5: Weighted Entropy for attribute

Entropy(A) = (3/10\*0.918) + (3/10\*0.918) + (4/10\*0.811) = 0.874

# Step 6: Information Gain

IG(S, checking\_status) = 0.971 - 0.874 = 0.097

# Step 7: Gini Index

Parent node: Gini(S) = 0.48

<0 DM: Gini = 0.444

0–200 DM: Gini = 0.444

no checking: Gini = 0.375

Weighted Gini(A) = 0.416

Gini Gain = 0.480 - 0.416 = 0.064

# Final Notes

- Information Gain (IG) ≈ 0.10

- Gini Gain ≈ 0.064

In building a Decision Tree, Weka’s J48 uses Information Gain Ratio, while CART uses Gini Index.