Decision Tree Practice Answers

NOTES FOR SHS | DATA SCIENCE | TOPIC 61

Worked mathematics

Parent Gini: 1 -(0.6 squared + 0.4 squared) = 0.48. Parent entropy: 0.9710 bits.

Left counts (5,1) yield Gini 0.2778; right (1,3) yield Gini 0.3750. Weight by 6/10 and 4/10 respectively: 0.3167. Reduction: 0.1633.

Weighted child entropy: 0.7145; parent minus child entropy = 0.2564 bits.

Practice solutions

1. Root = starting node; internal node = learned test; branches = outcomes; leaf = terminal prediction.

2. Gini for 8 A, 2 B is 1-(0.8 squared + 0.2 squared)=0.32.

3. Binary entropy for equal proportions is 1 bit.

4. Weighted impurity accounts for differing child sample sizes.

5. The test data would influence parameter choice, undermining an unbiased final evaluation.

6. Cost-complexity pruning penalizes large trees and removes branches when useful.

7. Macro-F1 averages class F1 equally; weighted-F1 weights classes by support.

8. Importance captures reduction under fitted splits and can favor flexible measurements; it is not a causal estimate.

9. Use permitted public benchmark records or a clearly synthetic dataset.

10. Load only self-created trusted joblib files in compatible environments because pickle-based files can execute code.