

Decision Tree Analysis

Nafis Nahian
CSE 318 - Artificial Intelligence Lab

July 1, 2025

1 Introduction

Decision trees are fundamental machine learning algorithms that create predictive models through recursive binary splits. The choice of splitting criterion significantly impacts tree performance, accuracy, and generalization capability. This study compares three splitting criteria:

- **Information Gain (IG)**: Measures reduction in entropy after splitting
- **Information Gain Ratio (IGR)**: Normalizes IG by split information to reduce bias toward multi-valued attributes
- **Normalized Weighted Information Gain (NWIG)**: A weighted variant considering attribute importance

2 Experimental Setup

2.1 Datasets

Two benchmark datasets were used:

- **Adult Dataset**: 32,562 records with binary classification (income greater than or less than 50K)
- **Iris Dataset**: 150 records with 3-class classification (iris species)

2.2 Methodology

- 20 independent runs with 80%-20% train-test split
- Maximum tree depths: 2, 3, and 4
- Metrics: Accuracy, training time, testing time, and total nodes

3 Results

3.1 Accuracy Analysis

3.1.1 Adult Dataset

Table 1 presents the accuracy results for the Adult dataset across different criteria and depths.

Table 1: Adult Dataset Accuracy Results (%)

Criteria	Depth 2	Depth 3	Depth 4
IG	81.98	82.37	82.05
IGR	83.36	83.35	86.02
NWIG	81.97	82.27	81.76

3.1.2 Iris Dataset

Table 2 shows the accuracy results for the Iris dataset.

Table 2: Iris Dataset Accuracy Results (%)

Criteria	Depth 2	Depth 3	Depth 4
IG	92.00	91.67	92.67
IGR	93.67	95.00	94.67
NWIG	92.00	91.67	92.67

3.2 Tree Complexity Analysis

3.2.1 Node Count Comparison

Table 3 compares the average total nodes across criteria and depths.

Table 3: Average Total Nodes Comparison

Criteria	Adult Dataset			Iris Dataset		
	D=2	D=3	D=4	D=2	D=3	D=4
IG	92	1,103	4,899	20	31	35
IGR	240	290	525	22	25	27
NWIG	127	1,307	5,278	20	30	34

3.3 Computational Performance

Table 4 summarizes the training and testing times for the Adult dataset.

Criteria	Training Time (s)		
	Depth=2	Depth=3	Depth=4
IG	0.64	3.41	40.94
IGR	1.15	1.36	2.04
NWIG	0.6	2.5	32.53

Table 4: Adult Dataset Computational Performance

4 Visualizations

4.1 Accuracy vs Depth Analysis

Figure 1 shows the accuracy comparison across different depths and criteria for both datasets.

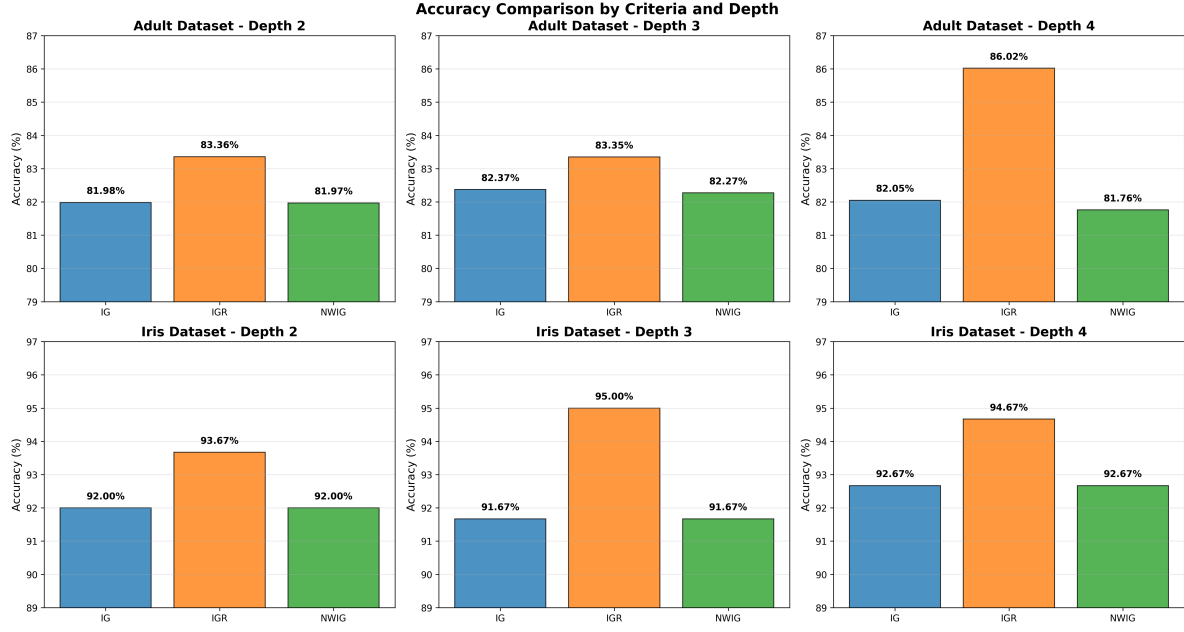


Figure 1: Accuracy Comparison by Criteria and Depth for Both Datasets

Figure 2 provides a direct comparison between datasets for each depth level.

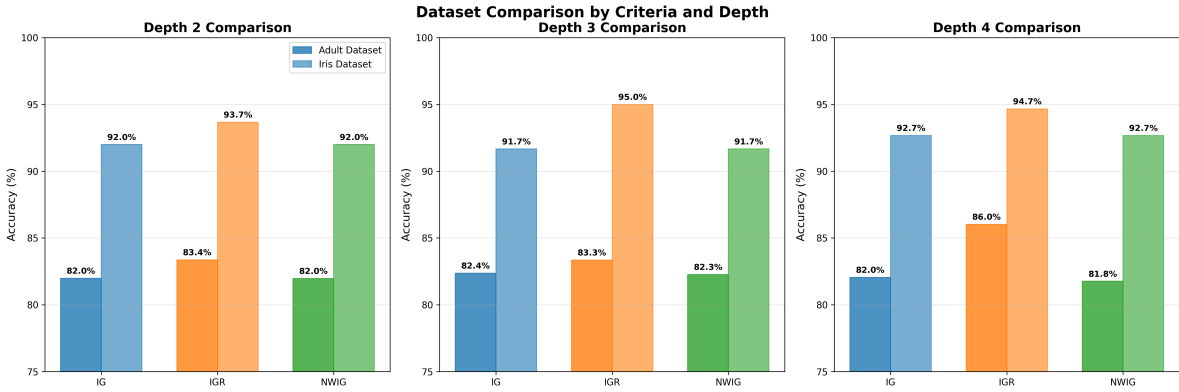


Figure 2: Dataset Comparison by Criteria and Depth

4.2 Tree Complexity Analysis

Figure 3 illustrates the relationship between tree depth and complexity measured by node count.

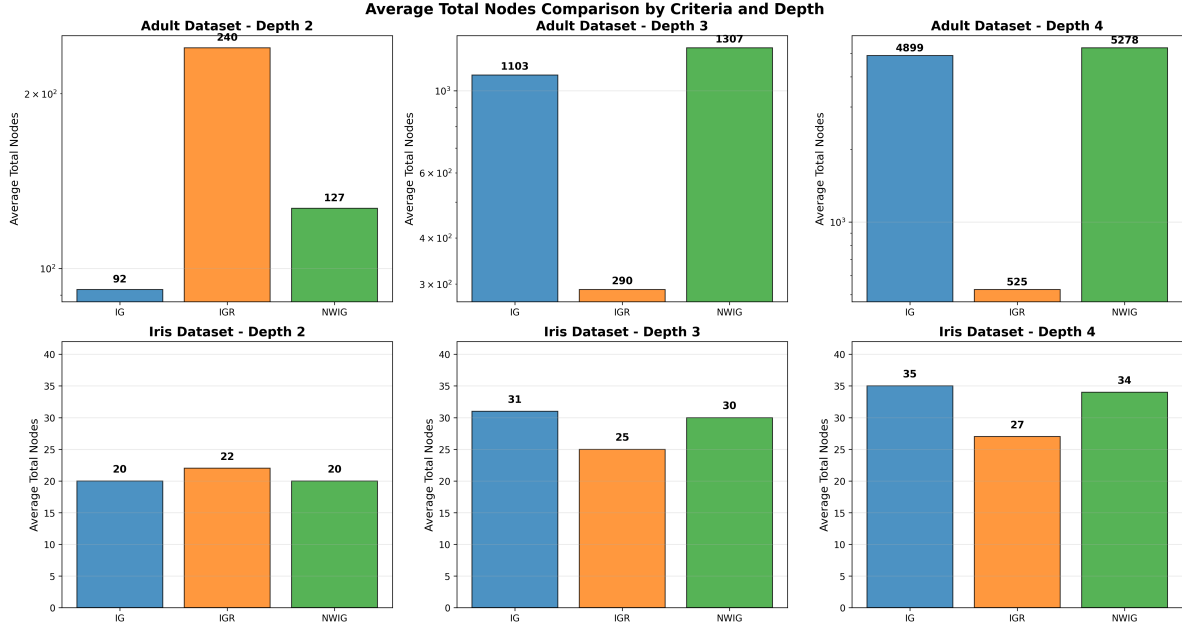


Figure 3: Average Total Nodes by Criteria and Depth

Figure 4 shows a direct comparison of node counts between datasets, highlighting the dramatic complexity differences.

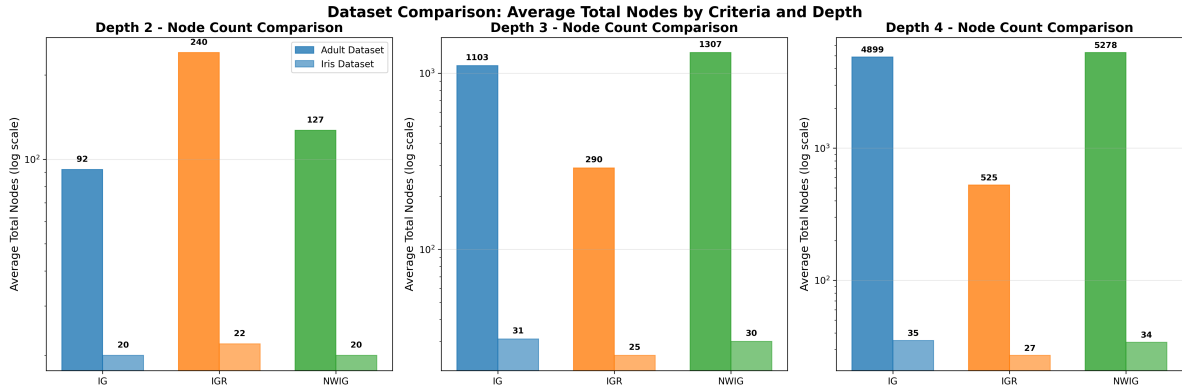


Figure 4: Node Count Comparison Between Datasets

5 Observations and Analysis

5.1 Key Findings

5.1.1 Information Gain Ratio (IGR) Superiority

- **Consistent Best Performance:** IGR achieved the highest accuracy across most depth-dataset combinations
- **Adult Dataset:** IGR peaked at 86.02% (depth 4), significantly outperforming IG (82.05%) and NWIG (81.76%)
- **Iris Dataset:** IGR achieved 95.00% accuracy at depth 3, the highest among all configurations

5.1.2 Overfitting Resistance

- **IG and NWIG Overfitting:** Both criteria showed performance degradation at depth 4 for the Adult dataset
- **IGR Robustness:** IGR continued improving with increased depth, suggesting better resistance to overfitting
- **Tree Complexity:** IGR produced significantly smaller trees (525 nodes vs. 4,899 for IG and 5,278 for NWIG at depth 4)

5.1.3 Computational Efficiency Trade-offs

- **Training Time:** IGR required moderate training time, avoiding the exponential growth seen in IG and NWIG
- **Testing Efficiency:** Despite larger initial testing times at shallow depths, IGR maintained efficient testing at depth 4
- **Scalability:** IGR's controlled node growth suggests better scalability for larger datasets

5.2 Dataset-Specific Patterns

5.2.1 Adult Dataset (Large, Complex)

- **Depth Sensitivity:** Significant performance variations with depth changes
- **Criterion Impact:** Choice of splitting criterion critically affects both accuracy and computational cost
- **Overfitting Evidence:** Clear overfitting patterns in IG and NWIG at maximum depth

5.2.2 Iris Dataset (Small, Simple)

- **High Baseline Performance:** All criteria achieved $\geq 90\%$ accuracy
- **Limited Depth Benefit:** Minimal improvement beyond depth 3
- **Criterion Robustness:** Less dramatic differences between criteria, though IGR still superior

6 Conclusions

6.1 Primary Recommendations

1. **Prefer IGR for Decision Tree Construction:** Consistently superior performance across datasets and depths
2. **Optimal Depth Selection:** Depth 3-4 provides best accuracy-complexity balance for most scenarios
3. **Dataset Considerations:** Simple datasets (like Iris) may not require deep trees, while complex datasets benefit from careful depth tuning

6.2 Theoretical Implications

- **Bias Reduction:** IGR's normalization effectively reduces bias toward multi-valued attributes
- **Overfitting Prevention:** The gain ratio mechanism provides implicit regularization
- **Scalability :** IGR's efficiency makes it suitable for larger, real-world applications