

## Decision Tree — Deep Dive for AI Engineers

### 1. What Is a Decision Tree?

A decision tree is a rule-based ML model that predicts outcomes by recursively splitting data using simple threshold-based questions.

### 2. Tree Structure

A tree consists of a root node, internal nodes, and leaf nodes where predictions are produced.

### 3. Decision Process

At each node, the tree evaluates all candidate splits and selects the one that best separates the data.

### 4. Regression Splits

For regression, splits minimize variance or mean squared error in child nodes.

### 5. Classification Splits

For classification, splits maximize class purity using Gini impurity or entropy.

### 6. Recursive Growth

The tree grows recursively until stopping criteria are met.

### 7. Strengths

Decision trees capture non-linear relationships, feature interactions, and require no feature scaling.

### 8. Overfitting

Trees can easily overfit by memorizing training data, leading to high variance.

### 9. Bias–Variance Tradeoff

Trees have low bias but high variance.

### 10. Regularization

Limiting depth, enforcing minimum samples per leaf, and minimum split sizes help control overfitting.

### 11. Practical Usage

Trees are rarely used alone in production due to instability.

### 12. Trees in Ensembles

Random Forest and boosting methods use trees as base learners.

### 13. Interpretability

Tree rules are human-readable and useful for explainability.

### 14. When to Use

Suitable for small datasets, explainability needs, and prototyping.

### 15. Limitations

Sensitive to noise, unstable, and cannot extrapolate trends.

**16. Interview Summary**

Decision trees recursively split data to minimize error but overfit easily, hence they are used inside ensembles.

**17. Key Takeaway**

Decision trees are foundational to modern ML but should rarely be deployed alone.