Decision Tree

**Interview Questions:**

1. What are some common hyperparameters of decision tree models, and how do they affect the model's performance?

**max\_depth**: Limits the maximum depth of the tree. A deeper tree can capture more complex patterns but may overfit the data.

**min\_samples\_split**: Minimum number of samples required to split an internal node. Higher values prevent overfitting by creating simpler trees.

**min\_samples\_leaf**: Minimum number of samples required to be at a leaf node. Helps in smoothing the model, especially with noisy data.

**max\_features**: The number of features to consider when looking for the best split. Reducing this can speed up training and reduce overfitting.

**max\_leaf\_nodes**: Limits the number of leaf nodes. Can control model complexity and prevent overfitting.

**criterion**: Function to measure the quality of a split (e.g., 'gini' for Gini impurity, 'entropy' for information gain). Affects how the tree makes decisions.

2. What is the difference between the Label encoding and One-hot encoding?

 **Label Encoding**:

* Converts categorical labels into numerical values (e.g., "Red" = 0, "Green" = 1, "Blue" = 2).
* Suitable for ordinal data (where order matters).
* May introduce unintended ordinal relationships in non-ordinal data, potentially misleading some models.

 **One-hot Encoding**:

* Converts categorical variables into a binary matrix (e.g., "Red" = [1,0,0], "Green" = [0,1,0], "Blue" = [0,0,1]).
* Works well for nominal data (no inherent order).
* Increases dimensionality, which can be inefficient with many unique categories.