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

Ans: Common decision tree hyperparameters and their effects:

1. **Max Depth**: Limits the tree's depth. Shallow trees may underfit, deep trees may overfit.
2. **Min Samples Split**: Minimum samples to split a node. Larger values reduce complexity and overfitting.
3. **Min Samples Leaf**: Minimum samples at a leaf node. Larger values prevent overfitting.
4. **Max Features**: Number of features to consider at each split. Fewer features reduce overfitting, but too few may hurt performance.
5. **Criterion**: Function to measure split quality. "Gini" and "entropy" are common choices.
6. **Max Leaf Nodes**: Limits leaf nodes, reducing complexity and overfitting.
7. **Splitter**: Strategy for splitting nodes. "Best" is optimal but slower; "Random" is faster but may reduce accuracy.
8. **Max Samples**: For ensemble models, limits training data per tree to reduce variance and overfitting.

Adjusting these parameters balances model complexity, preventing overfitting or underfitting.

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

Ans : The key differences between **Label Encoding** and **One-Hot Encoding** are:

1. **Representation**:
    a. **Label Encoding**: Converts categorical values into integer labels (e.g., "Red" → 0, "Green" → 1, "Blue" → 2).
    b. **One-Hot Encoding**: Creates binary columns for each category (e.g., "Red" → [1, 0, 0], "Green" → [0, 1, 0], "Blue" → [0, 0, 1]).
2. **Use Case**:
    a. **Label Encoding**: Useful when there is an inherent ordinal relationship between categories (e.g., "Low", "Medium", "High").
    b. **One-Hot Encoding**: Suitable for nominal categories with no inherent order (e.g., "Red", "Green", "Blue").
3. **Dimensionality**:
    a. **Label Encoding**: Results in a single column of integers, thus not increasing the feature space.
    b. **One-Hot Encoding**: Increases the feature space, creating as many columns as there are unique categories.
4. **Interpretability**:

a. **Label Encoding**: Can introduce unwanted ordinal relationships if used on nominal data (e.g., "Red" being coded as 0, "Green" as 1, and "Blue" as 2 could imply an arbitrary ordering).
b. **One-Hot Encoding**: Avoids introducing any unintended relationships by treating each category independently.

In summary, **Label Encoding** is simpler but better suited for ordinal data, while **One-Hot Encoding** is more appropriate for categorical data with no intrinsic order.