**Decision Tree - Interview Questions and Answers**

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

Decision tree models have several hyperparameters that can significantly impact their performance.

**Some common hyperparameters include:**

**Max Depth:**

This hyperparameter controls the maximum depth of the decision tree.

Increasing the max depth can lead to overfitting, while decreasing it can lead to underfitting.

**Min Samples Split:**

This hyperparameter specifies the minimum number of samples required to split an internal node.

Increasing this value can reduce overfitting, while decreasing it can lead to overfitting.

**Min Samples Leaf:**

This hyperparameter specifies the minimum number of samples required to be at a leaf node.

Increasing this value can reduce overfitting, while decreasing it can lead to overfitting.

**Max Features:**

This hyperparameter specifies the maximum number of features to consider when looking for the best split.

Increasing this value can lead to overfitting, while decreasing it can lead to underfitting.

**Criterion:**

This hyperparameter specifies the function to measure the quality of a split.

Common criteria include Gini impurity and entropy.

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

Label encoding and one-hot encoding are two common techniques used to convert categorical variables into numerical variables that machine learning algorithms can process.

**Label Encoding:**

Label encoding is a simple technique that assigns a unique integer value to each category in a categorical variable.

**For example**, if we have a categorical variable "color" with three categories:

"red", "green", and "blue"

we can assign the values 0, 1, and 2 to each category.

The problem with label encoding is that it can create a false sense of hierarchy or ordering between the categories.

In the above example, the model may assume that "green" is somehow "better" or "higher" than "red", which is not necessarily true.

**One-Hot Encoding:**

One-hot encoding is a technique that creates a new binary feature for each category in a categorical variable.

**For example**, if we have a categorical variable "color" with three categories:

"red", "green", and "blue"

we can create three new binary features.

In this representation, each category is represented by a binary vector (a vector with only 0s and 1s) where all the elements are 0 except for the one corresponding to the category, which is 1.

**One-hot encoding is a more robust technique than label encoding because it does not create a false sense of hierarchy or ordering between the categories.**