Q1 What are some common hyperparameters of decision tree models and how do they affect the models performance ?

In decision tree models, there are several hyperparameters that significantly influence the model's performance.

1. Maximum Depth (max\_depth)

* Description: Controls the maximum depth of the tree.

2. Minimum Samples per Split (min\_samples\_split)

* Description: Specifies the minimum number of samples required to split an internal node.

3. Minimum Samples per Leaf (min\_samples\_leaf)

* Description: Defines the minimum number of samples that must be present in a leaf node.

4. Maximum Features (max\_features)

* Description: Specifies the number of features to consider when looking for the best split.

5. Maximum Leaf Nodes (max\_leaf\_nodes)

* Description: Limits the number of leaf nodes in the tree.

Q2 What is the diference between the label encoding and one-hot encoding ?

| **Aspect** | **Label Encoding** | **One-Hot Encoding** |
| --- | --- | --- |
| **How it works** | Converts each category into a unique integer value. | Creates a new binary column for each category. |
| **Output** | A single column with integer values. | Multiple columns with binary (0 or 1) values. |
| **Example** | Color: ["red", "blue", "green"] | Color: ["red", "blue", "green"] |
| **Encoded Output** | red → 0, blue → 1, green → 2 | red → [1, 0, 0], blue → [0, 1, 0], green → [0, 0, 1] |
| **Advantages** | Simple, efficient, reduces dimensionality. | No ordinal relationship between categories. |
| **Disadvantages** | Can introduce unintended ordinal relationships. | Increases dimensionality (especially for many categories). |
| **Use Case** | When the categorical feature has an ordinal relationship. | When categories are nominal (no intrinsic order). |