# **Decision Tree**

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

In decision trees, there are a few key hyperparameters that really impact how the model performs

* **max\_depth:** This limits how deep the tree can go. If the tree is too deep, it might overfit the training data. On the other hand, if it’s too shallow, it might underfit and miss important patterns.
* **min\_samples\_split:** This controls the minimum number of samples needed to split a node. Increasing this value makes the tree more conservative and helps prevent overfitting.
* **min\_samples\_leaf:** Similar to the one above, but it applies to the leaf nodes. It ensures that a leaf has a certain number of samples, which smooths the model a bit.
* **max\_features:** This limits the number of features the model considers when looking for the best split. It can be useful for regularization and adds randomness in ensemble models like Random Forests.
* **criterion:** This decides how the quality of a split is measured — typically either Gini impurity or Information Gain (entropy). Both works well, and the choice sometimes comes down to preference or experimentation.

So, tuning these hyperparameters helps control the complexity of the tree and balance between overfitting and underfitting.

**2. What’s the difference between Label Encoding and One-Hot Encoding?**

Both are techniques for converting categorical data into numbers, but they’re used differently:

Label Encoding assigns each category a number. So, if you have colours like Red, Green, and Blue, it might turn them into 0, 1, and 2. It’s simple and works well when the categories have a natural order — like "Low", "Medium", "High".

One-Hot Encoding, on the other hand, creates a new column for each category. For example, if you have three colours, it’ll create three columns: one for Red, one for Green, and one for Blue. Only one column will have a 1 (true) for each row, and the rest will be 0.

Label encoding is faster and uses less memory, but it can be misleading if the categories don’t actually have an order. One-hot encoding avoids that issue but can lead to high dimensionality if you have a lot of unique categories.