Decision Trees

What is a Decision Tree?

- A Decision tree is a supervised machine learning algorithm.
- It is generally used when the relationship between X and Y are non linear.
- It splits the data into various possible entities concerning a specific parameter.

Terminologies of a Decision Tree?

- **Root Node:** This is where the decision tree starts. The root node represents the complete dataset which divides into two or more comparable sets.
- **Leaf Node:** The leaf nodes represent the final output. The algorithm can not segregate the tree further after coming to the leaf node.
- **Splitting:** This term means dividing the root node or the decision node into different sub-nodes according to the underlying conditions.

Types of Decision Tree?

DT are of two types -

- **Classification Tree:** In this the leaf nodes have a discrete value
- **Regression Tree :** In this the leaf nodes have continuous real values

Advantages of a Decision Tree?

- DT are easy to understand and use.
- The outcomes of DT are purely based on the decisions.
- The DT can handle both numerical and categorical features with minimal preprocessing.
- DT does not need the features to be scaled or normalised.

Disadvantages of a Decision Tree?

- DT are prone to overfitting.
  - approaches like early stopping and pruning reduces overfitting
- If some class dominate the data, then the final tree is biased towards that class.
- It is complicated to choose the perfect attribute to split a node.