**Q-1: What is Pruning in Decision Trees ? Why is it important ?**

Pruning a decision tree is a fundamental step in optimizing the computational efficiency as well as classification accuracy. Pruning usually results in reducing size of tree, avoids unnecessary complexity, and to avoid overfitting of the data sets when classifying new data.

Overfitting can lead to an excessively large number of rules, many of which have very little predictive value for unseen data.

**Post-pruning**

Post-pruning is also known as backward pruning. In this, first Generate the decision tree and then remove non-significant branches. Post-pruning a decision tree implies that we begin by generating the (complete) tree and then adjust it with the aim of improving the classification accuracy on unseen

instances. There are two principal methods of doing this. One method that is widely used begins by converting the tree to an equivalent set of rules. Another commonly used approach aims to retain the decision tree but to replace some of its subtrees by leaf nodes, thus converting a complete tree to a

smaller pruned one which predicts the classification of unseen instances at least as accurately. There are various methods for the post pruning.

1. Reduced Error Pruning

2. Error complexity pruning

3. Minimum Error pruning

4. Cost based pruning

**Pre-pruning**

Pre-pruning is also called forward pruning or online-pruning.Pre-pruning prevent the generation of non-significant branches. Pre-pruning a decision tree involves using a ‘termination condition’ to decide when it is desirable to terminate some of the branches prematurely as the tree is generated. When constructing the tree some significant measures can be used to assess the goodness of a split. If

partitioning the tuples at a node would result the split that falls below a prespecified threshold, then further partitioning of the given subset is halted otherwise it is expanded. High threshold result in oversimplified trees, whereas low threshold result in very little simplification. There are various approaches for the pre-pruning.

1. Minimum no of object pruning

2. Chi-square pruning

**Why is it important ?**

1. Pruning is important as it produces statistical measurements.

2. Accuracy and tree length are measured for various pruning factors for large data sets. When no of fold increases, accuracy almost increases and size of tree sometimes reduced.

3. Helps in optimizing computational efficiency as well as classification accuracy.

4. Although the decision tree generated by the ID3 are accurate and efficient, but they often provide very large trees that make them incomprehensible. When decision tree induced, many of the branches will reflect anomalies in the training data due to noise. This problem of overfitting happens when the learning algorithm continues to develop hypotheses that reduce training set error at the cost of an increased test set errors. To overcome these problem of overfitting, pruning is necessary.