**Unit-2**

**Chapter-2.1**

**Lecture 2.1.3**

CO-3,4,5

PRUNING IN DECISION TREE

**Pruning**

The performance of a tree can be further increased by ***pruning***. *It involves****removing the branches that make use of features having low importance***. This way, we reduce the complexity of the tree, and thus increasing its predictive power by reducing overfitting.

Pruning can start at either root or the leaves. The simplest method of pruning starts at leaves and removes each node with the most popular class in that leaf, this change is kept if it doesn't deteriorate accuracy. Its also called **reduced error pruning**. More sophisticated pruning methods can be used such as **cost complexity pruning** where a learning parameter (alpha) is used to weigh whether nodes can be removed based on the size of the sub-tree. This is also known as **weakest link pruning.**

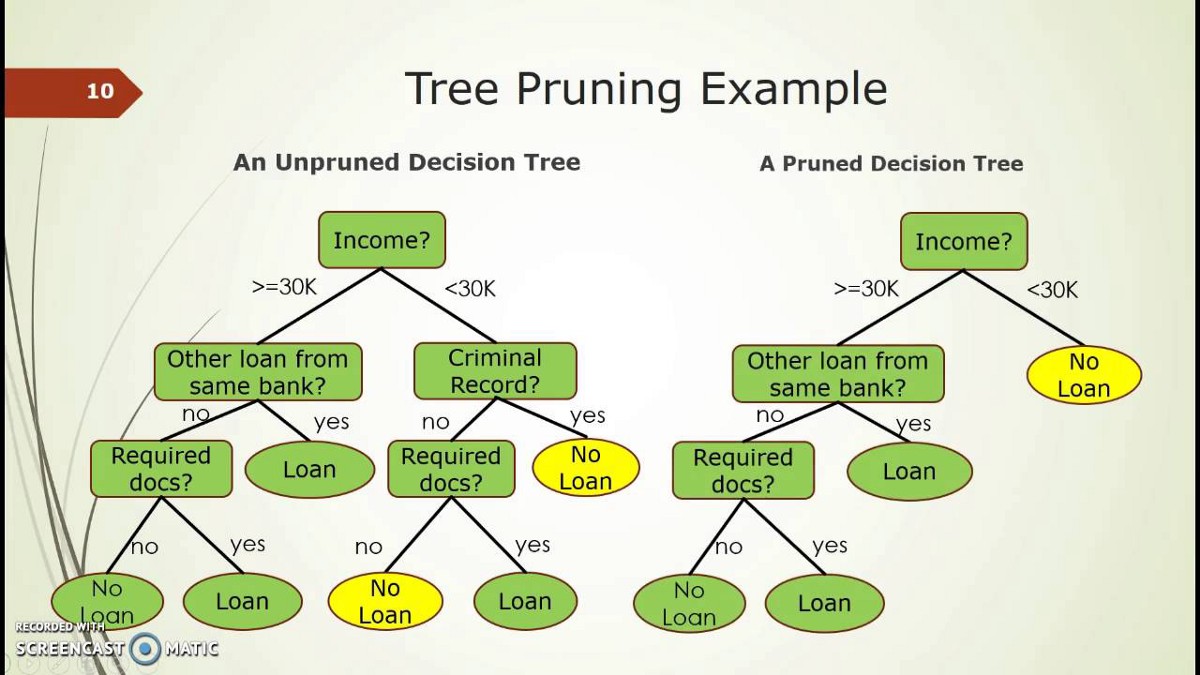


Fig-1 Pruning

* There is another technique for reducing the number of attributes used in a tree - *pruning*
* Two types of pruning:
* Pre-pruning (forward pruning)
* Post-pruning (backward pruning)

In prepruning, we decide during the building process when to stop adding attributes (possibly based on their information gain)

* However, this may be problematic – Why?
* Sometimes attributes individually do not contribute much to a decision, but combined, they may have a significant impact

Postpruning waits until the full decision tree has built and then prunes the attributes

* Two techniques:
* Subtree Replacement
* Subtree Raising

**Book Reading and Video Material**

* Understanding Machine Learning: From Theory to Algorithms by Shai Shalev-Shwartz and Shai Ben-David-Cambridge University Press 2014 [Download](https://www.cse.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf) Buy at Amazon
* Introduction to Machine Learning – the Wikipedia guide
* [Pruning](https://youtu.be/t56Nid85Thg)

**Pruning.pptx**