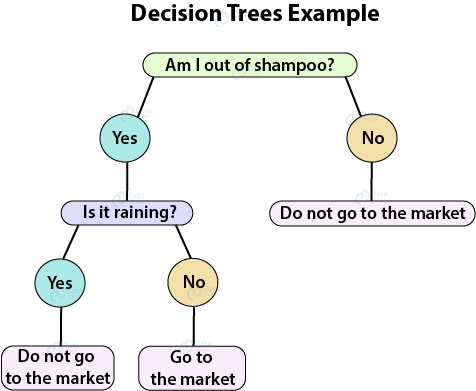
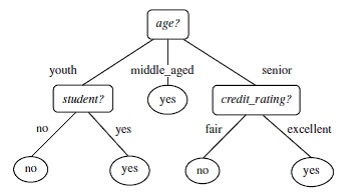
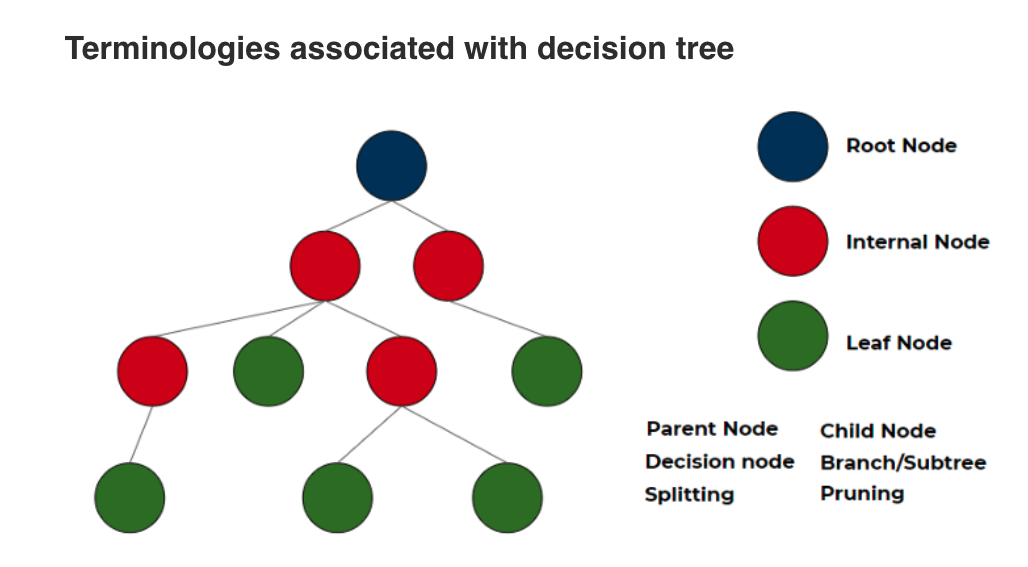
Decision Tree

A decision tree is a diagram or chart that people use to determine a course of action or show a statistical probability. We use decision tree in daily life. For example;



Another example;





**Parent node:** In any two connected nodes, the one which is higher hierarchically, is a parent node.

**Child node:** In any two connected nodes, the one which is lower hierarchically, is a child node.

**Root node:** The starting node from which the tree starts, it has only child nodes. The root node does not have a parent node.

**Leaf Nodes:** Nodes at the end of the tree, which do not have any children are leaf nodes or called simply leaf.

**Internal Nodes:** All the in-between the root node and the leaf nodes are internal nodes or simply nodes. Internal nodes have both a parent and at least one child.

**Splitting:** Dividing a node into two or more sub-nodes or adding two or more children to a node.

**Decision node:** When a parent splits into two or more children nodes then that node is called a decision node.

**Pruning:** When we remove the sub-node of a decision node, it is called pruning. You can understand it as the opposite process of splitting.

**Branch/Sub-tree:** a subsection of the entire tree is called a branch or sub-tree.

The decision of splits heavily affects a tree’s accuracy. There are multiple algorithms for splitting. Gini impurity, Chi-Square, information gain, reduction in variance.

We will talk about gini impurity and information gain.

Complete later

<https://www.analyticsvidhya.com/blog/2020/10/all-about-decision-tree-from-scratch-with-python-implementation/>

Advantages

Easy to visualize and interpret

Useful in data exploration

Less data cleaning required

The data type is not a constraint

Disadvantages

Overfitting

Not exact fit for continuous data

Python Sklearn Implementation

Import libraries

Load data

Separate futures and target

(Split data into train and test for some cases to find accuracy, Also we may need to normalize input for some ML algorithm)

Define the model

Fit the Model

Predict