Decision Tree

Supervised Machine Learning
Classification/Regression Algorithm

Course Topics

- ✔ What is Decision Tree ?
- ✔ Decision Tree Terminology
- ✔ How Decision Tree works?
- ✔ Entropy and information gain
- ✔ Pros and Cons of Decision Tree
- ✔ Applications of Decision Tree
- ✔ Optimizing Decision Tree model performance
- ✔ Modeling Decision Tree with Python

SKILLOGIC

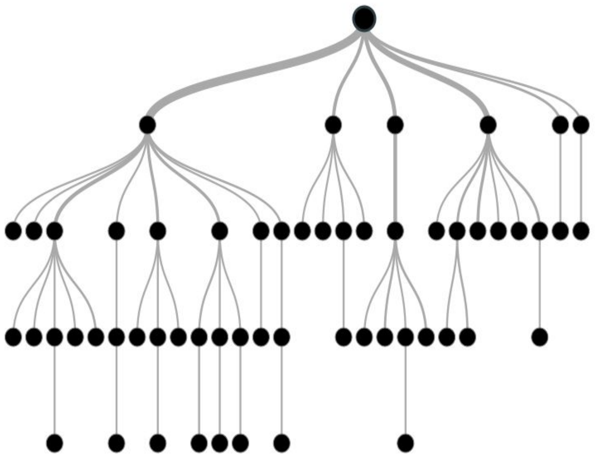
Datamites

What is Decision Tree ?

- Decision tree is a supervised machine learning algorithms.

- It is use for both classification and regression problem.

- Decision tree work on simple nested if else statement condition.



SKILLOGIC

Types of Decision Trees

Types of decision tree is based on the type of target variable we have. It can be of two types:

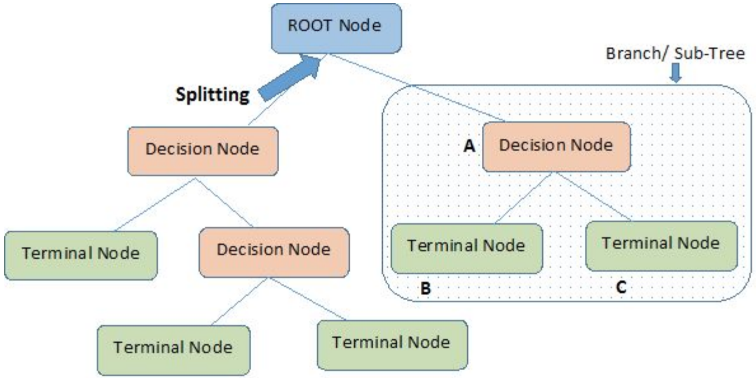
- **Categorical Variable Decision Tree:** Decision Tree which has a categorical variable as target.
- **Continuous Variable Decision Tree:** Decision Tree which has a continuous variable as target.

SKILLOGIC

Datamites

Decision Tree Terminology

- **Root Node:** It represents entire population or sample and this further gets divided into two or more homogeneous sets.
- **Splitting:** It is a process of dividing a node into two or more sub-nodes.
- **Decision Node:** When a sub-node splits into further sub-nodes, then it is called decision node.

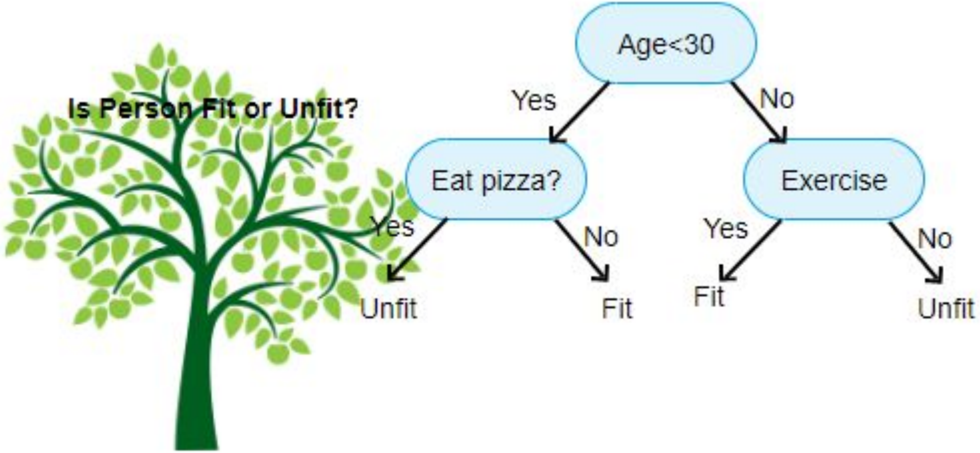


Note:- A is parent node of B and C.

- **Leaf/ Terminal Node:** Nodes with no children (no further split) is called Leaf or Terminal node.
- **Pruning:** When we reduce the size of decision trees by removing nodes (opposite of Splitting), the process is called pruning.
- **Branch / Sub-Tree:** A subsection of decision tree is called branch or sub-tree.
- **Parent and Child Node:** A node, which is divided into sub-nodes is called parent node, where sub-nodes are the child of parent node.

How Decision Tree works?

Decision tree work on simple nested if else statement condition. Decision tree break the dataset into smaller and smaller subset while at the same time it create a decision tree and result determine by leaf nodes.



SKILLOGIC

How Decision Tree works?

Let's understand with examples:-

Predict if John will play tennis

Training examples: **9 yes / 5 no**

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

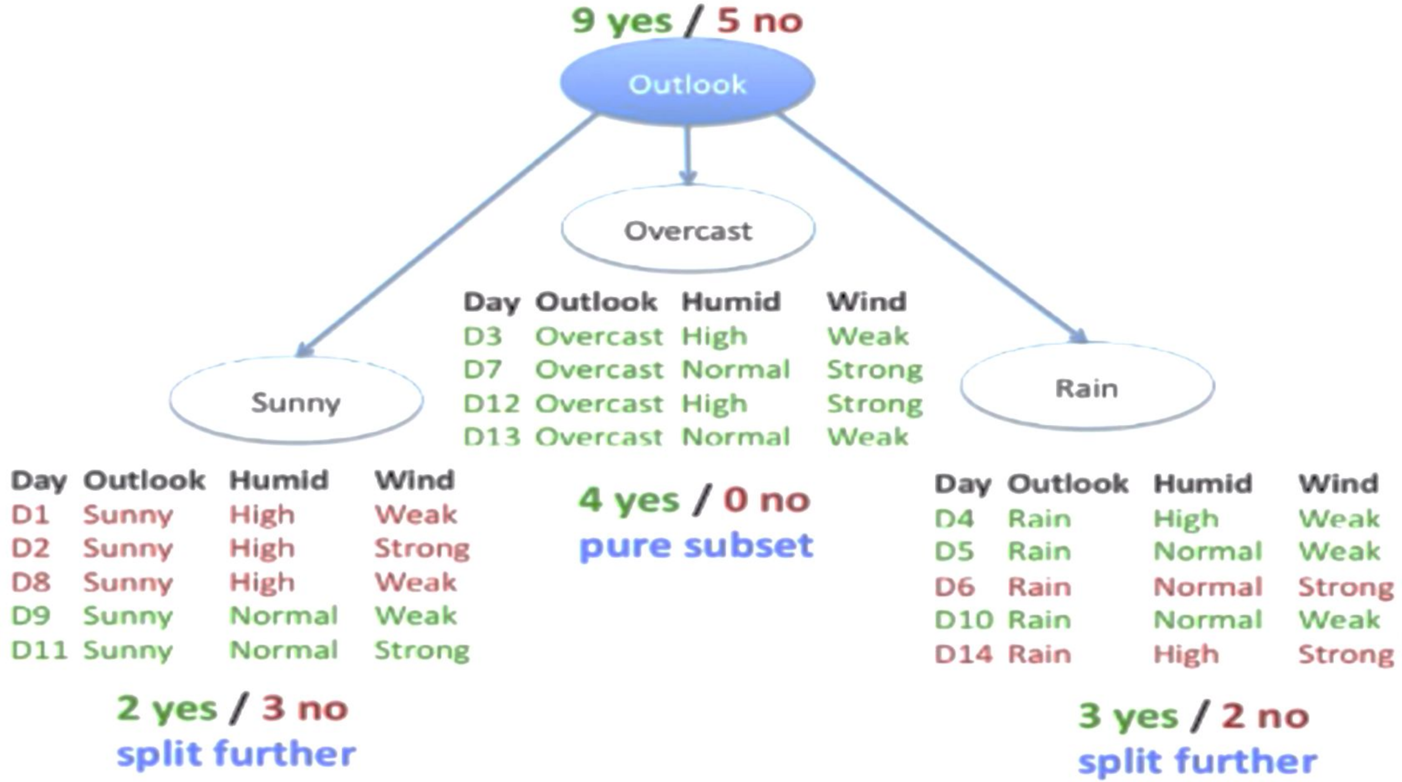
New data:

Day	Outlook	Humidity	Wind	Play
D15	Rain	High	Weak	?

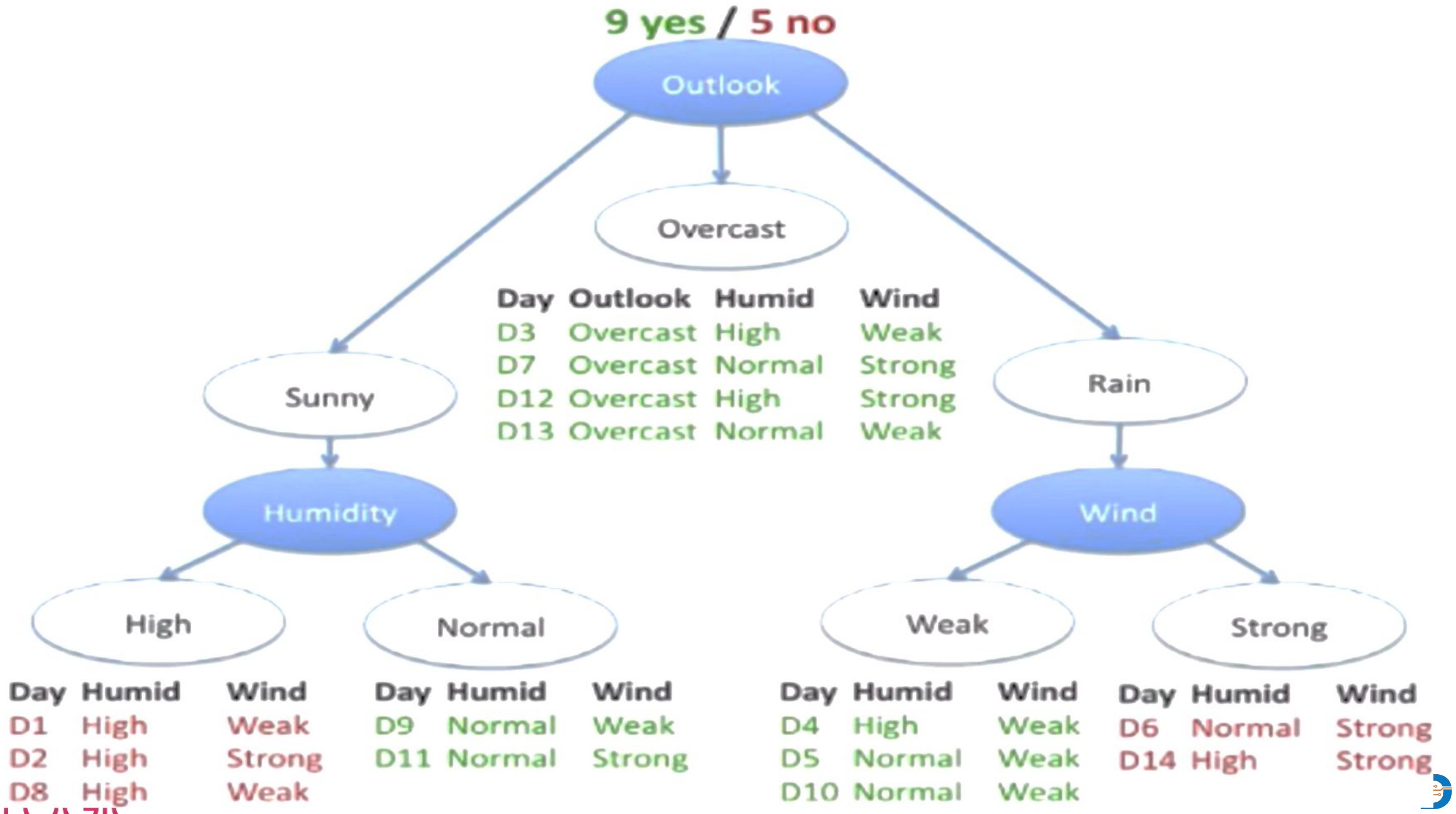
Datamites

How Decision Tree works?

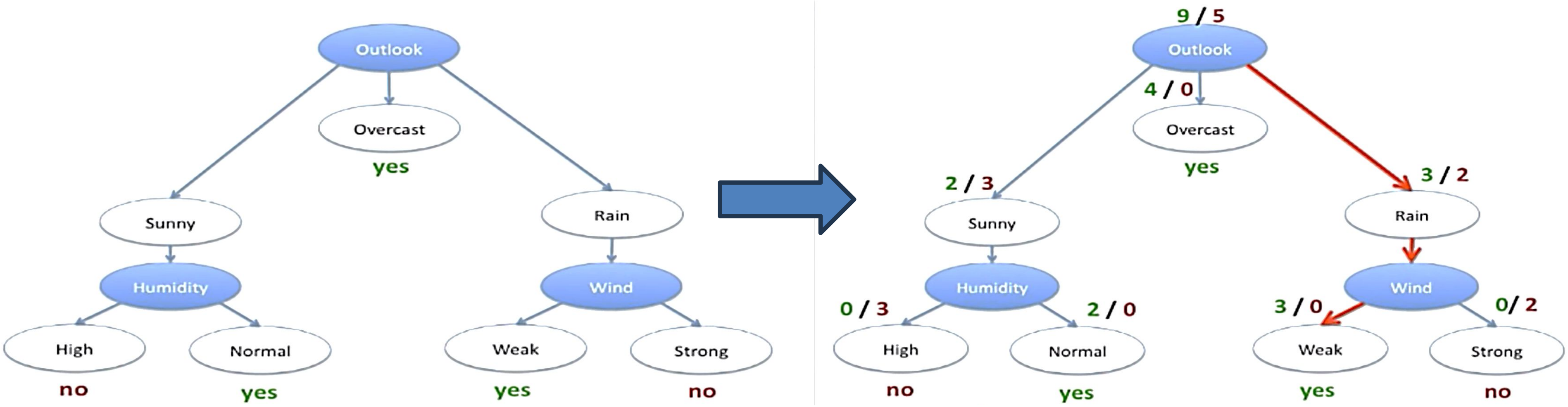
Let’s see what is the probability of yes and no.



How Decision Tree works?



How Decision Tree works?



New data:

Day	Outlook	Humid	Wind	
D15	Rain	High	Weak	→ Yes

How to select the right feature?

ENTROPY MEASURES HOMOGENEITY OF EXAMPLES

- Entropy measures the *impurity* of a collection of examples. It depends from the distribution of the random variable p.
  - S is a collection of training examples
  - p_+ the proportion of positive examples in S
  - p_- the proportion of negative examples in S

$$Entropy(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

Examples

$$Entropy([14+, 0-]) = -14/14 \; log_2 (14/14) - 0 \; log_2 (0) = 0$$

$$Entropy([9+, 5-]) = -9/14 \; log_2 (9/14) - 5/14 \; log_2 (5/14) = 0.94$$

$$Entropy([7+, 7-]) = -7/14 \; log_2 (7/14) - 7/14 \; log_2 (7/14) = 1/2 + 1/2 = 1$$

Datamites

Information Gain

Information Gain:

Information Gain is used to determine which feature/attribute gives us the maximum information about a class.

- Information Gain is based on the concept of entropy, which is the degree of uncertainty, impurity or disorder.

- Information Gain aims to reduce the level of entropy starting from the root node to the leave nodes.

- The greater the reduction in the uncertainty, the more information is gained about Y from X.

Entropy and Information Gain

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$\begin{aligned}
Values(Wind) &= Weak, Strong \\
S &= [9+, 5-] \\
S_{Weak} &\leftarrow [6+, 2-] \\
S_{Strong} &\leftarrow [3+, 3-] \\
Gain(S, Wind) &= Entropy(S) - \sum_{v \in \{Weak, Strong\}} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= Entropy(S) - (8/14) Entropy(S_{Weak}) \\
&\quad - (6/14) Entropy(S_{Strong}) \\
&= 0.940 - (8/14)0.811 - (6/14)1.00 \\
&= 0.048
\end{aligned}$$

GINI INDEX

GINI INDEX

Gini index or Gini impurity measures the degree or probability of a particular variable.

$$\text{Gini Index} = 1 - \sum_{i=1}^{n} (P_i)^2$$

4 records 2 yes, 2 No

- Probability of +ve class = 2/4
- Probability of –ve class = 2/4

Gini index = 1 – [sq(p+) + sq(p-)] = 1 – [sq(2/4) + sq(2/4)] = 0.50

Pros of Decision Tree

- Easy to Understand: Decision tree output is very easy to understand even for people from non-analytical background. It does not require any statistical knowledge to read and interpret them.
- Less data cleaning required: It requires less data cleaning compared to some other modeling techniques. It is not influenced by outliers and missing values to a fair degree.
- Data type is not a constraint: It can handle both numerical and categorical variables.
- Decision Tree Versatility : Decision trees can be customized for a variety for situations.

Cons of Decision Tree

- Over fit : Over fitting is one of the most practical difficulty for decision tree models.
- Low accuracy for continuous variables: While working with continuous numerical variables, decision tree loses information when it categorizes variables in different categories.
- They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree.

Applications of Decision Tree

Modeling Decision Tree

Decision Tree in action