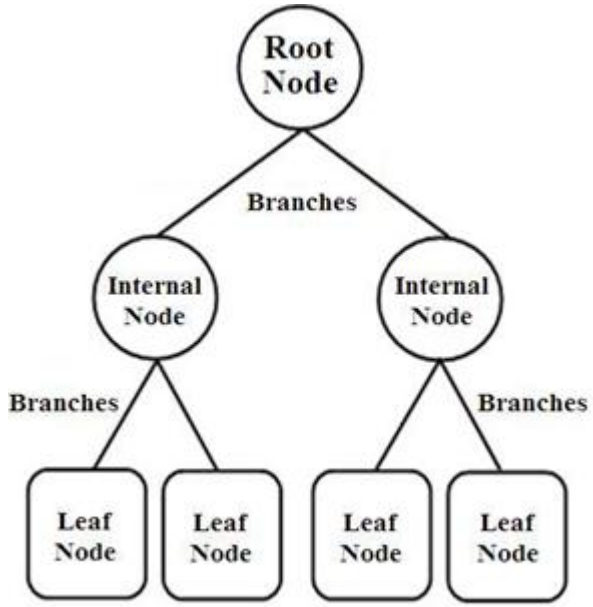
Decision Tree

What is a tree?

A tree has nodes and branches.

A rooted tree has Root node and leafs.

- **Rooted node** - have children.
- **Leafs**-do not have any children.



What is a decision tree?

- It is a tree-structured classifier.
- Decision trees can be used both for classification and regression.
- It has two types of nodes.
  - Decision Node
  - Leaf Node

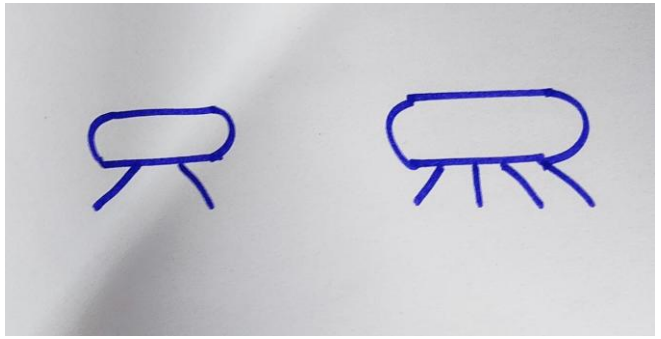
Decision Node & Leaf Node

Decision Node :

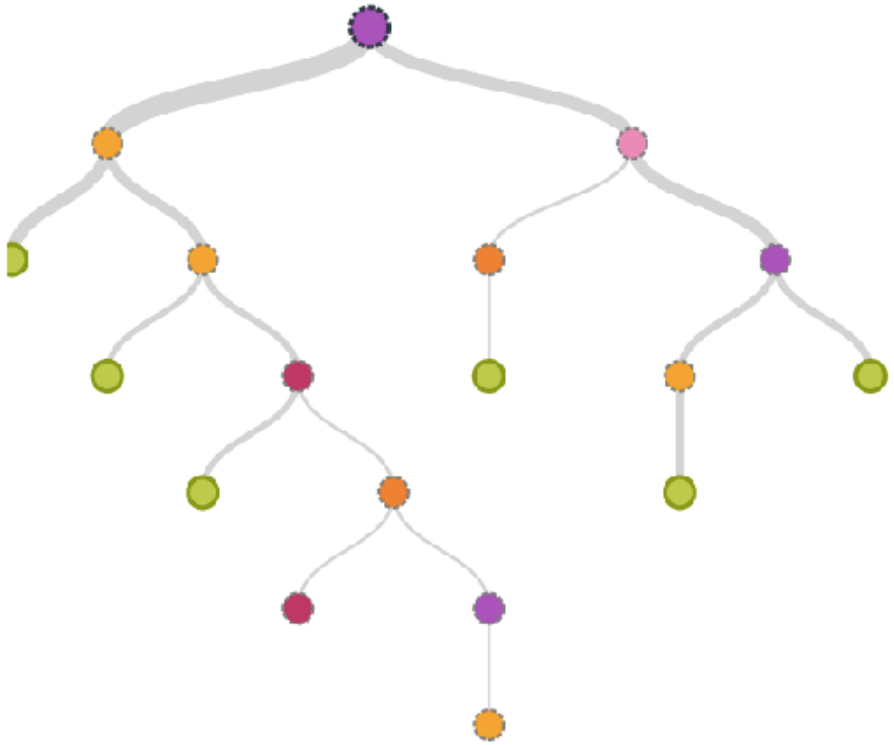
- In decision nodes - specify a choice or a test.
- Based on this you can decide which direction you can go.
- This test is usually done on the value of a feature or attribute of the instance.

Leaf node:

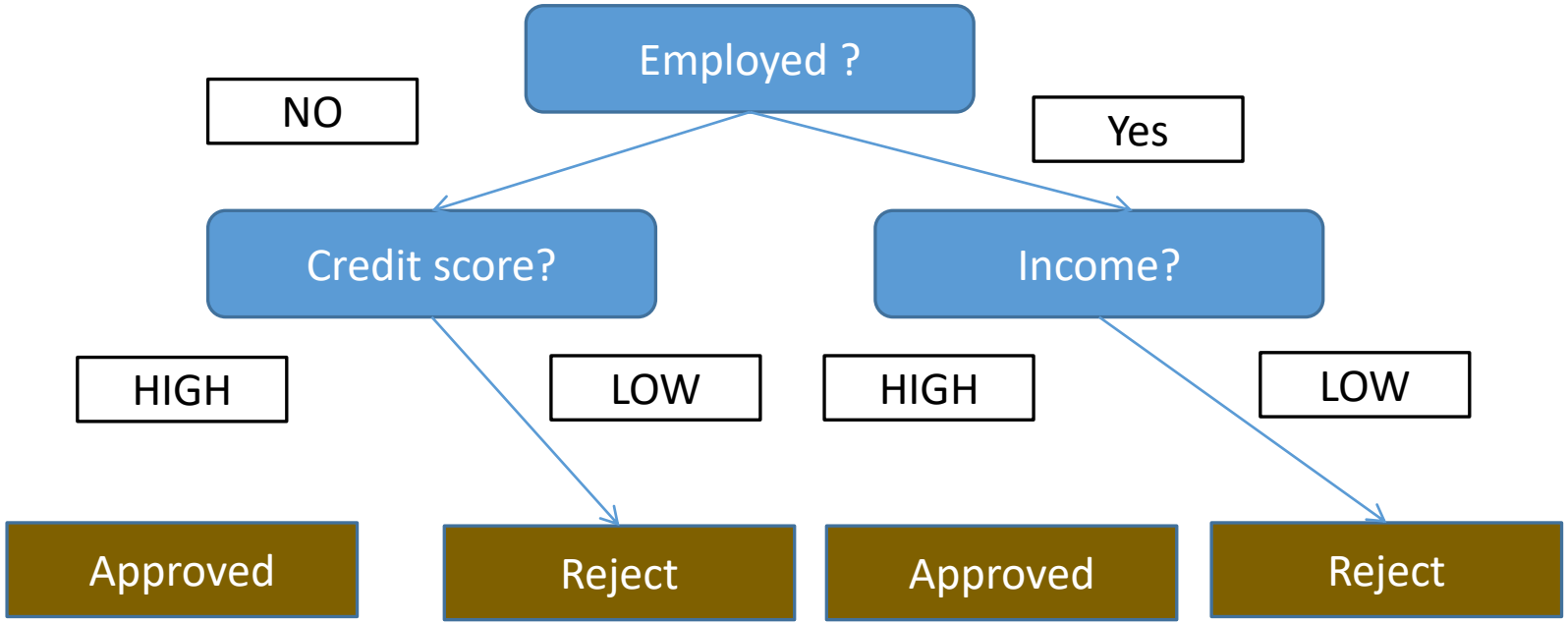
- leaf node indicate the classification of an example or the value of the example.



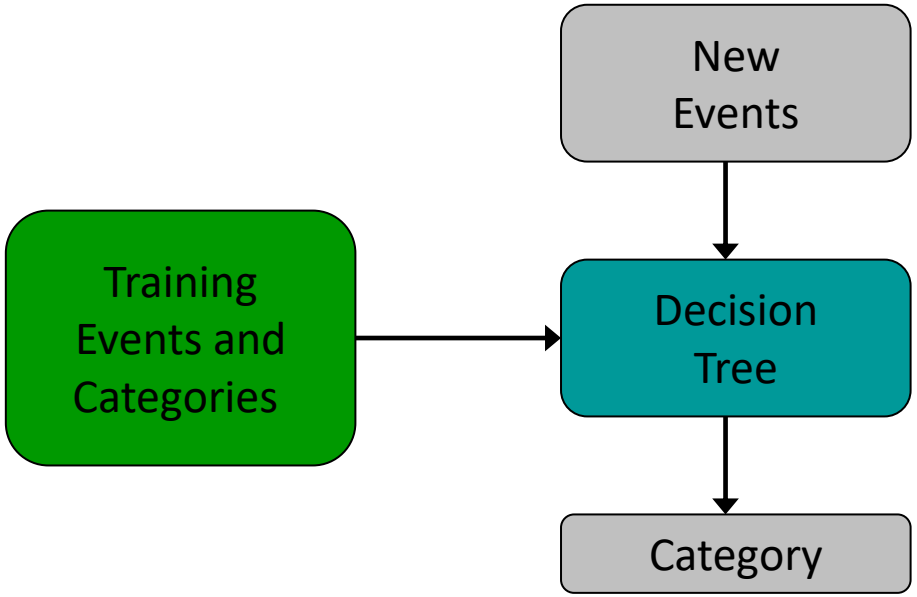
Example



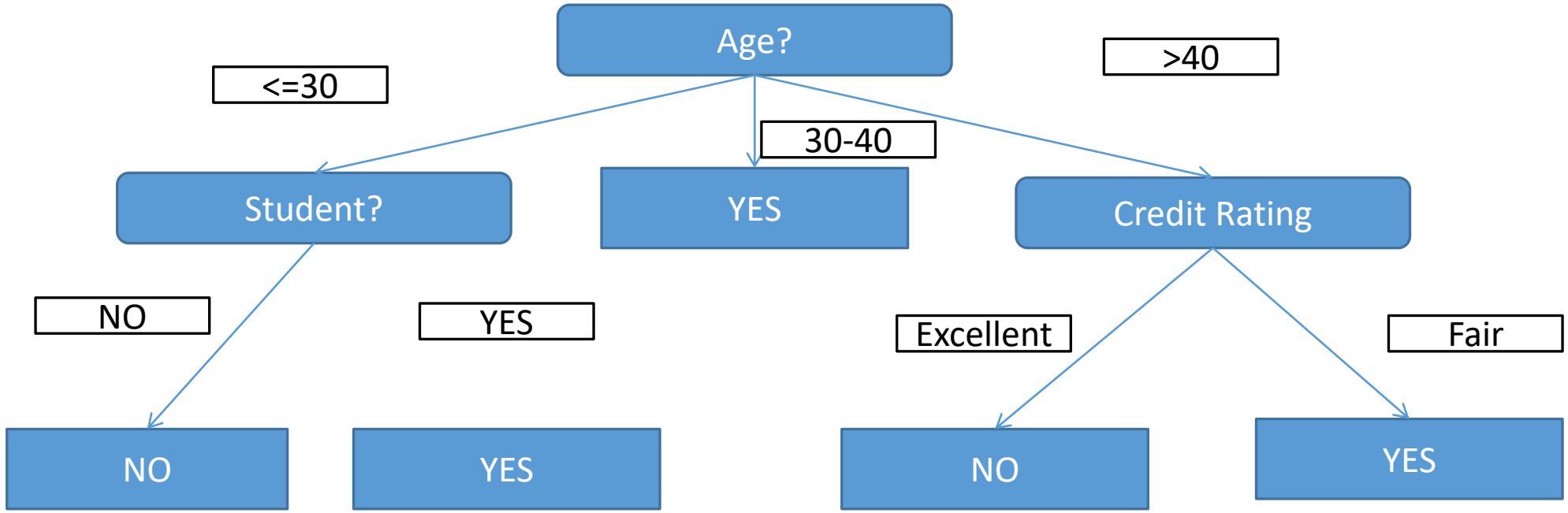
Example 1: Draw a decision tree about whether to approve a loan.



- Use a decision tree to predict categories for new events.
- Use training data to build the decision tree.



Example 2: Decision tree to decide whether a person is likely to buy a computer.

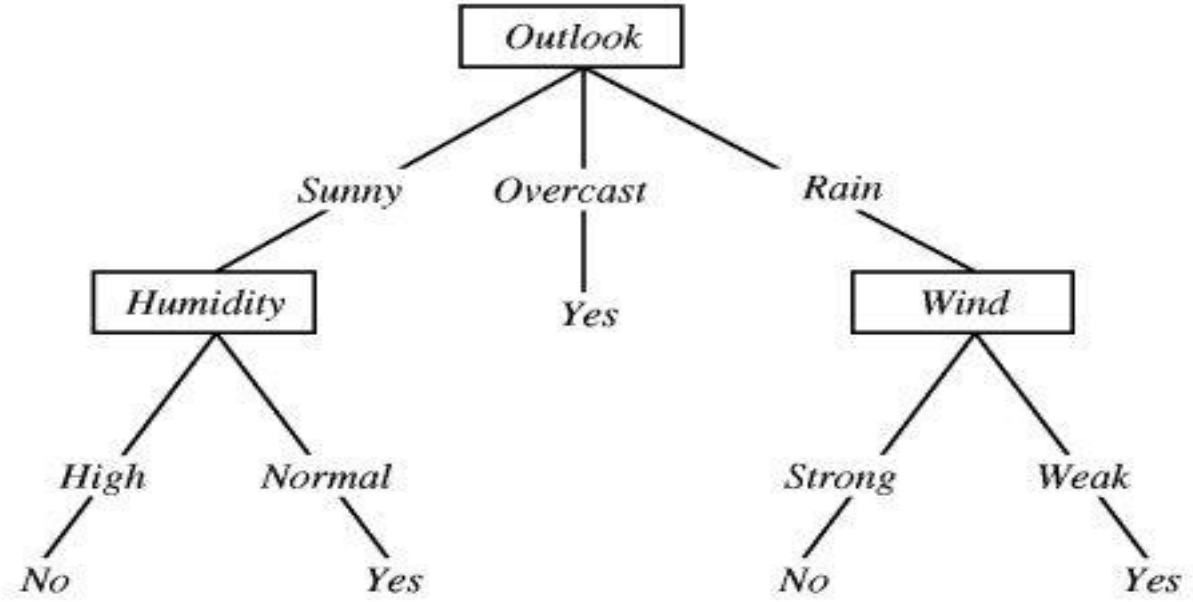


Example 3: Decision tree to decide whether it is a good day to play tennis

The attributes used in the decision tree are :

- Outlook - sunny, overcast or rainy
- Humidity- high and normal;
- Wind - strong and weak;
- Temperature - hot, mild and cool

Target concept : play tennis, yes or no



- Each internal node tests an attribute
- Each branch corresponds to attribute value
- Each leaf node assigns a classification

ID3 in brief

- ID3 stands for Iterative Dichotomiser 3 and is named such because the algorithm iteratively (repeatedly) dichotomizes(divides) features into two or more groups at each step.

- Developed by Ross Quinlan, ID3 uses a **top-down greedy** approach to build a decision tree.

- In simple words, the **top-down** approach means that we start building the tree from the **top**, and the **greedy** approach means that at each iteration, we select the best feature at the present moment to create a node.

- Most generally, ID3 is only used for classification problems with **nominal** features.