University of Roehampton London

Machine Learning

Dr Changjiang He, Dr Kuo-Ming Chao
Computer Science| School of Art
University of Roehampton

University of Roehampton London

Lesson 3.2

Decision Tree

Decision Tree

University of Roehampton London

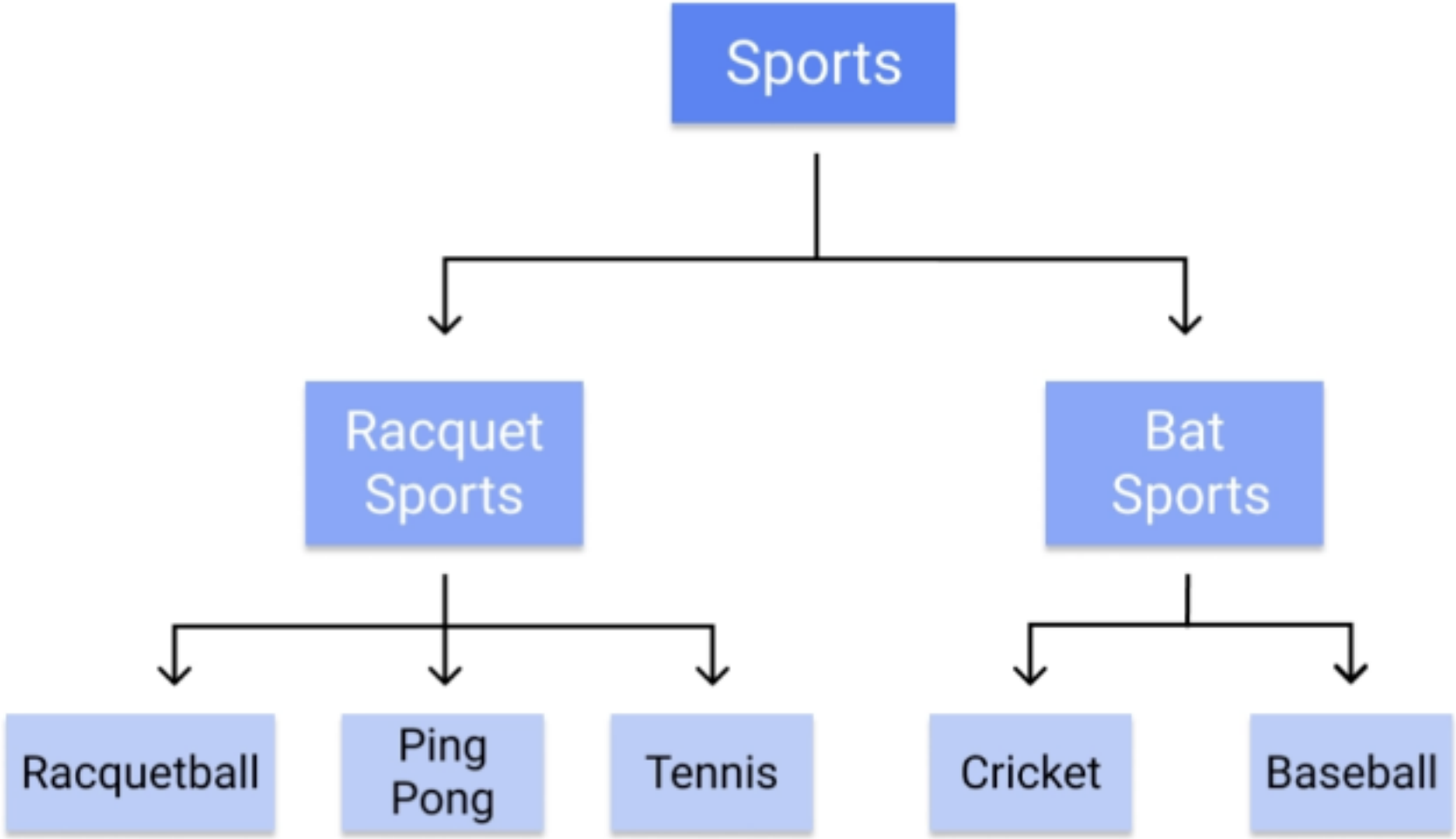
- A decision tree is a supervised learning algorithm that is perfect for classification problems, as it's able to order classes on a precise level.

- It works like a flow chart, separating data points into two similar categories at a time from the "tree trunk" to "branches," to "leaves," where the categories become more finitely similar.

- This creates categories within categories, allowing for organic classification with limited human supervision.

Decision Tree

University of Roehampton London



Terminologies

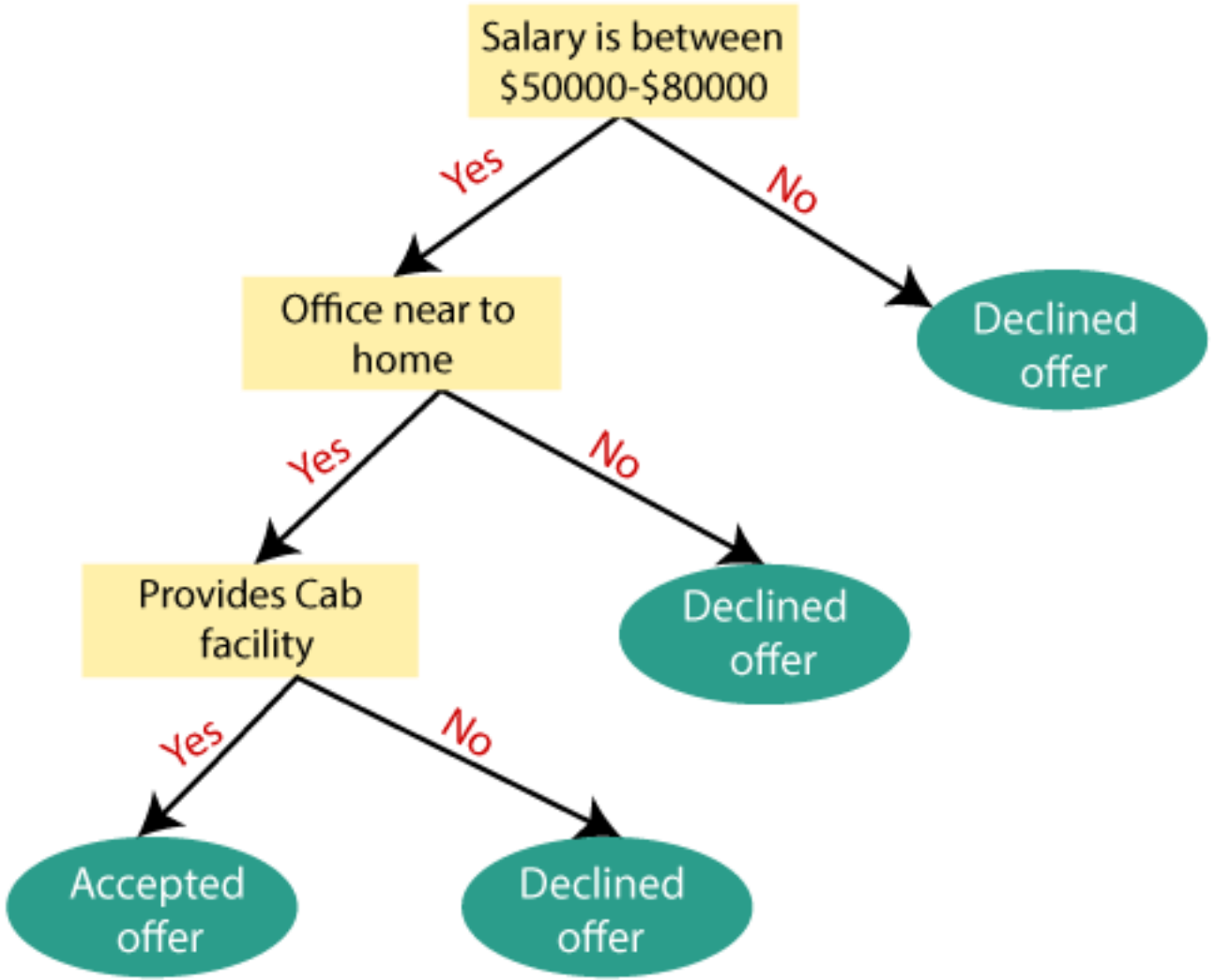
University of Roehampton London

- **Root Node:** Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.
- **Leaf Node:** Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
- **Splitting:** Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
- **Branch/Sub Tree:** A tree formed by splitting the tree.
- **Pruning:** Pruning is the process of removing the unwanted branches from the tree.
- **Parent/Child node:** The root node of the tree is called the parent node, and other nodes are called the child nodes.

Example

University of Roehampton London

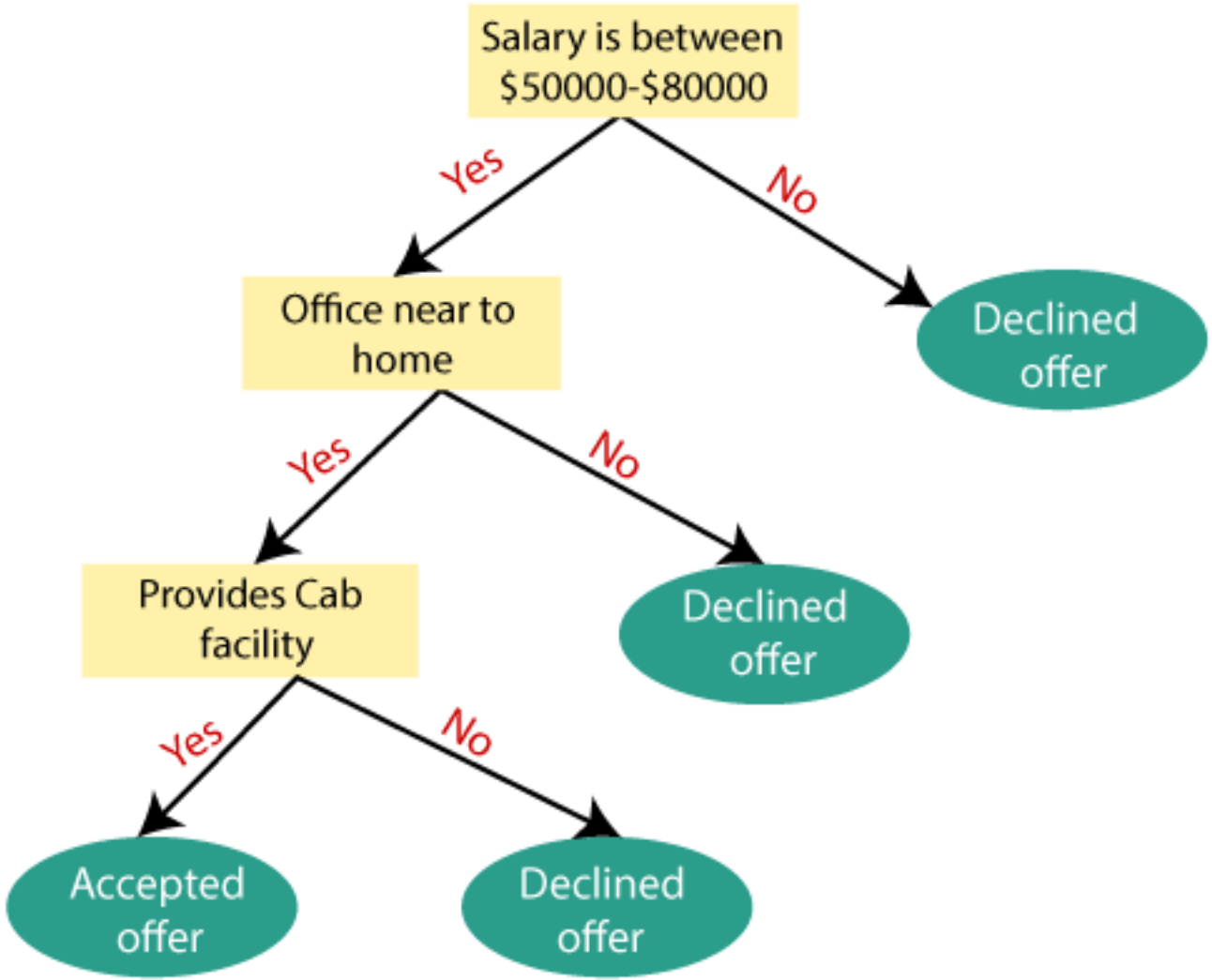
- Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not.
- We can derived a decision tree as shown in the figure.



Example

University of Roehampton London

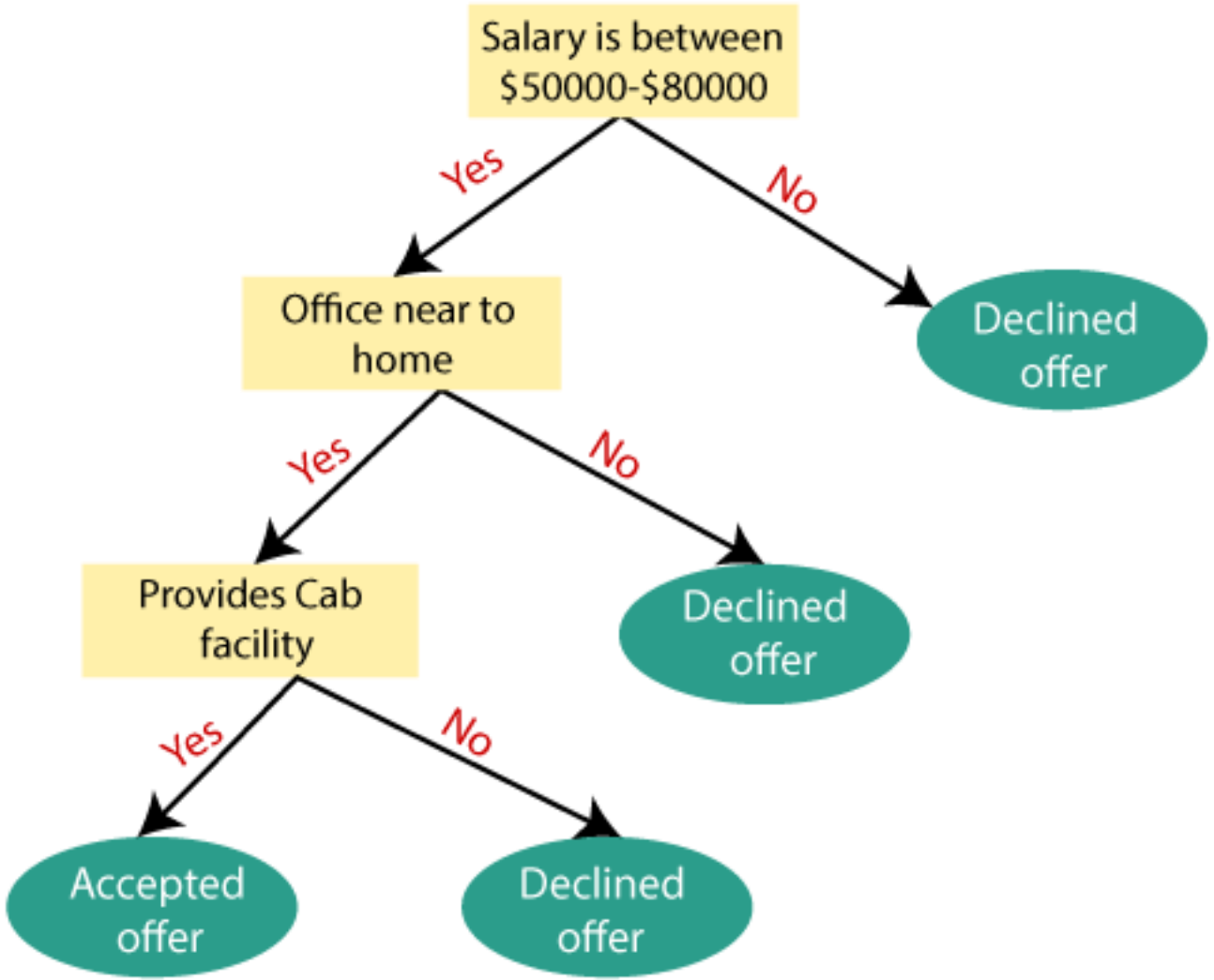
- So, to solve this problem, the decision tree starts with the root node (Salary attribute).



Example

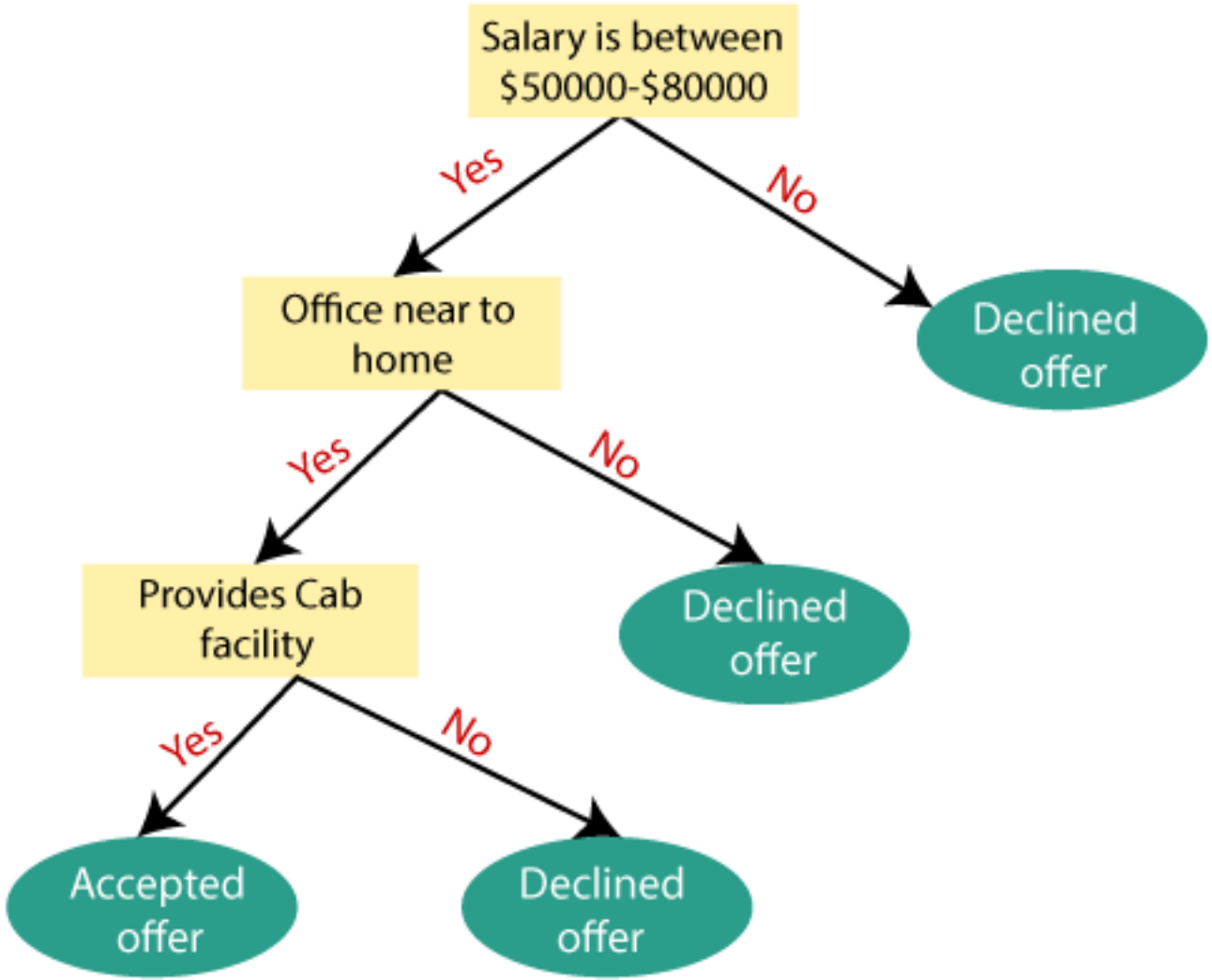
University of Roehampton London

- The root node splits further into the next decision node (distance from the office) and one leaf node based on the corresponding labels.



Example

- The next decision node further gets split into one decision node (Cab facility) and one leaf node.
- Finally, the decision node splits into two leaf nodes (Accepted offers and Declined offer).



How does the Decision Tree algorithm Work?

University of Roehampton London

- In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree.

- This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node.

- For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree. The complete process can be better understood using the below algorithm:

A Summary of the Process

University of Roehampton London

- **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.
- **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM).**
- **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
- **Step-4:** Generate the decision tree node, which contains the best attribute.
- **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.