Future Connect Media

Machine learning Part-A

Part of Future Connect Media's IT Course

By Abhishek Sharma

Future Connect Media

Topics to be covered:

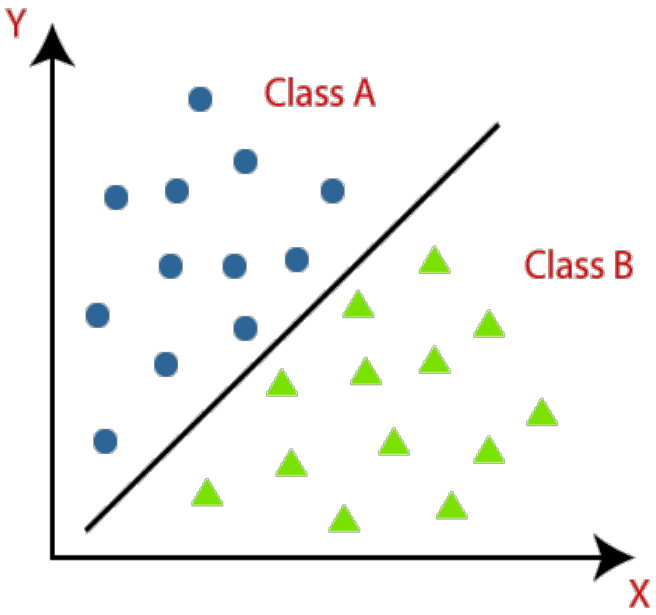
- Classification
- Model for Classification
- Decision Tree Classification
- Impurities in a Decision Tree

Classification

- A classification problem is used to identify specific categories of new observations based on one or more independent variables
- The main goal of the Classification algorithm is to identify the category of a given dataset, and these algorithms are mainly used to predict the output for the categorical data.

Some applications are:

- Image classification
- Fraud Detection
- Document classification
- Spam Filtering



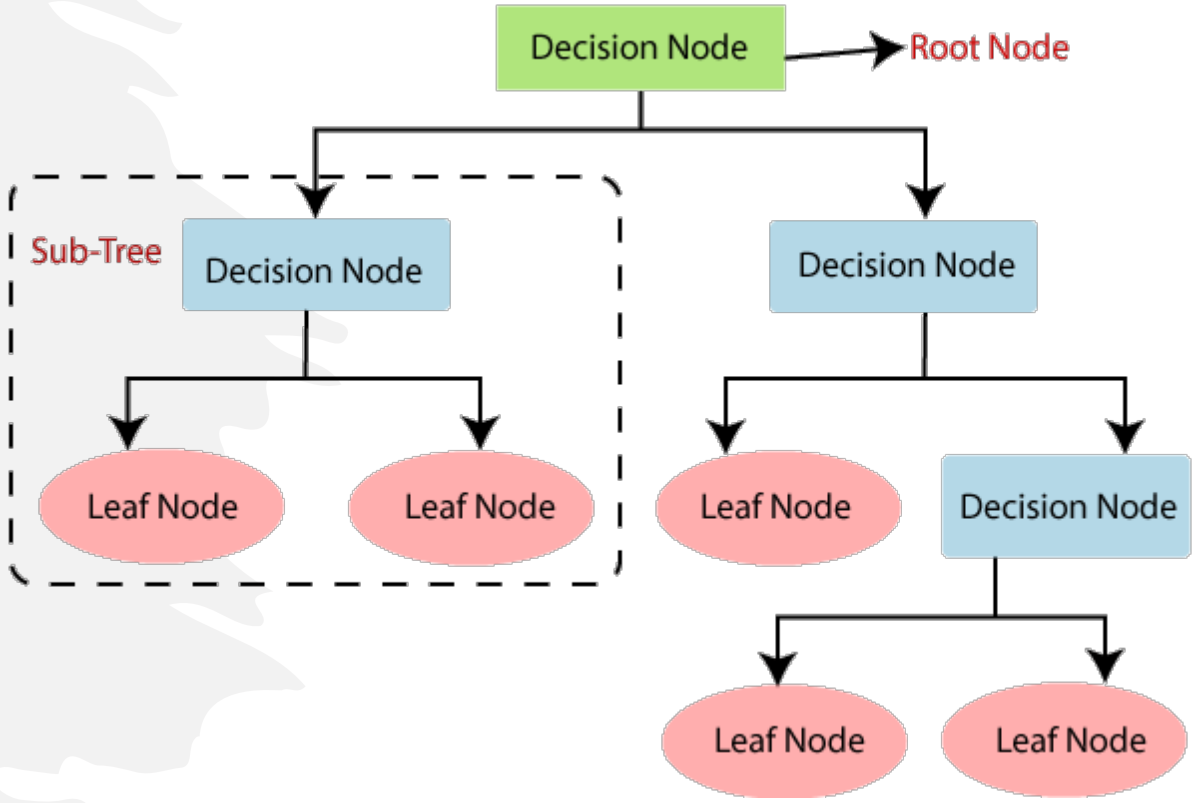
Models for Classification

- Following are some Supervised Machine learning models used for Classification:

  - Decision Tree Classification
  - Random Forest Classification
  - Logistic Regression
  - Support Vector Machine
  - Kernal SVM
  - K-Nearest neighbor
  - Naïve Bayes

Decision Tress Classification

- **Decision Trees (DTs)** are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.



Future Connect Media

Terminologies of Decision Tree

Future Connect Media

- Root Node (Root)
- Internal Node (Node)
- Branches
- Leaves

Impurity in Decision Tree

Leaves that contain mixture of classification are called **Impure**.

Methods for quantification of Impurities:

1. Gini Impurity
2. Entropy
3. Information Gain