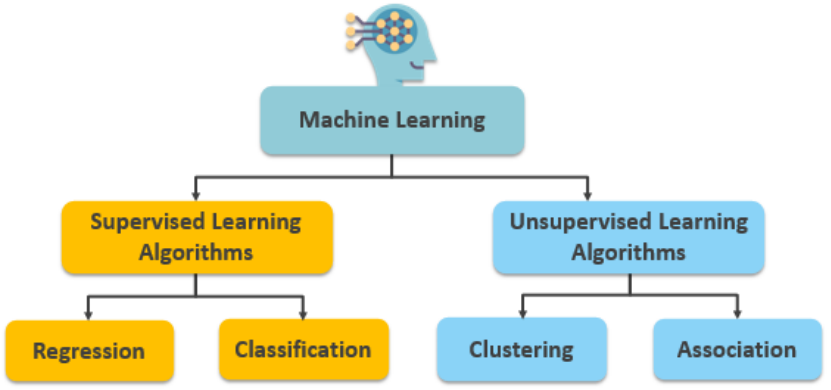
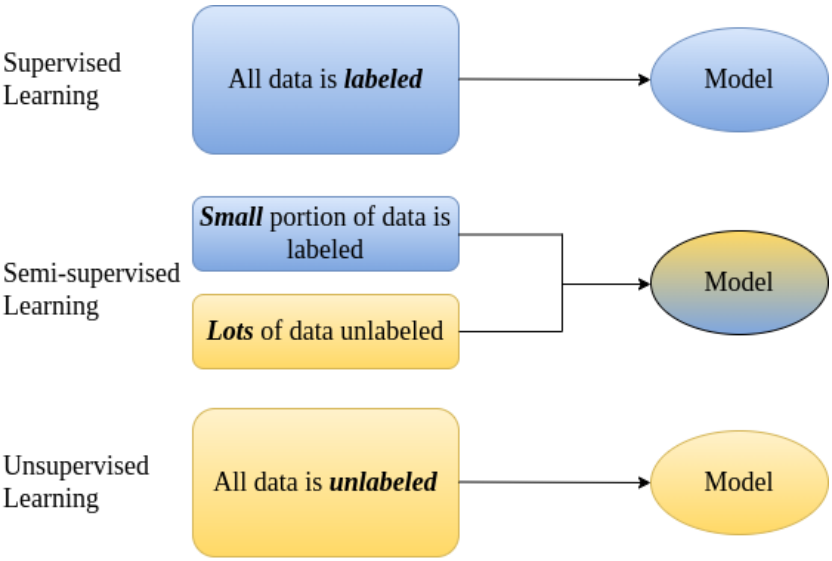
Unit-II
Classification





The following are the classification Algorithms:

- kNN
- Naïve Bayes Classifier
- Support Vector Machines
- Decision Tree

K-Nearest Neighbour(KNN)

K-Nearest Neighbour (KNN) is a supervised machine learning algorithm used for classification and regression tasks. It predicts the class or value of a data point based on its nearest neighbours in the feature space. KNN is a lazy learning algorithm, meaning it doesn't build a model during training but instead uses the entire dataset to make predictions at the time of testing.

How KNN Works

1. Choose the number of neighbours (k):
   Select the number of closest data points (neighbours) to consider for making predictions.
2. Calculate the distance:
   Compute the distance between the new data point and all the points in the dataset.
   Common distance metrics include Euclidean, Manhattan and Minkowski distances.
3. Select the nearest neighbours:
   Identify the k closest data points to the new data point.
4. Vote or average (based on task):
   - For classification, assign the class based on the majority vote of the neighbours.
   - For regression, predict the output as the average value of the neighbours.

Naïve Bayes Classifier

Naïve Bayes is a supervised machine learning algorithm based on Bayes' Theorem, commonly used for classification tasks. It assumes that features are independent given the class label, which simplifies calculations and makes it highly efficient.

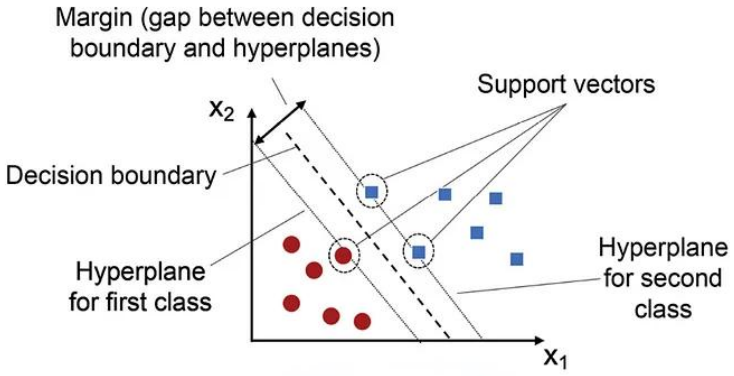
Formula of Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where:

- $P(A|B)$ → **Posterior probability** (Probability of event A occurring given B is true)
- $P(B|A)$ → **Likelihood** (Probability of observing B given A has occurred)
- $P(A)$ → **Prior probability** (Initial probability of event A before observing B)
- $P(B)$ → **Evidence** (Overall probability of event B)

Support Vector Machines (SVM): Support Vector Machine (SVM) is a supervised learning algorithm used for classification and regression tasks. It is particularly effective for high-dimensional data and works well for both linear and non-linear classification problems.



Decision Tree: A Decision Tree is a supervised learning algorithm used for classification and regression tasks. It works by breaking down a dataset into smaller subsets while forming a tree-like structure with decision nodes and leaf nodes.

A **Decision Tree** follows a **tree-like structure**:

- **Root Node**: The first splitting point based on the most significant feature.
- **Decision Nodes**: Intermediate nodes where data is split based on conditions.
- **Leaf Nodes**: Terminal nodes when no further splitting is possible.