**SUPERVISED LEARNING:**

[Supervised learning](https://www.ibm.com/think/topics/supervised-learning) is a machine learning approach that’s defined by its use of labeled data sets. These data sets are designed to train or “supervise” algorithms into classifying data or predicting outcomes accurately. Using labeled inputs and outputs, the model can measure its accuracy and learn over time.

Supervised learning can be separated into two types of problems when [data mining](https://www.ibm.com/think/topics/data-mining): classification and regression:

* **Classification** problems use an algorithm to accurately assign test data into specific categories. Or, in the real world, supervised learning algorithms can be used to classify spam in a separate folder from your inbox. Linear classifiers, support vector machines, decision trees and [random forest](https://www.ibm.com/think/topics/random-forest) are all common types of classification algorithms.
* **Regression** is another type of supervised learning method that uses an algorithm to understand the relationship between dependent and independent variables. Regression models are helpful for predicting numerical values based on different data points, such as sales revenue projections for a given business. Some popular regression algorithms are linear regression, logistic regression, and polynomial regression.

Classification projects:

1. Fake news detection
2. Spam email filtering
3. Customer churn prediction
4. Heart disease prediction
5. Image classification

Regression projects:

1. Stock price prediction
2. House price prediction
3. Loan approval prediction
4. Sales forecasting
5. Crop yield prediction

Classification Algorithm:

1. Naive Bayes: Spam detection
2. Linear Discriminant Analysis (LDA): Face recognition
3. k-Nearest Neighbors (KNN) : Image classification

Regression Algorithm:

1. Linear Regression: House price prediction
2. Elastic Net : Stock price prediction
3. Support Vector Regression (SVR): Time series forecasting

**UNSUPERVISED LEARNING:**

[Unsupervised learning](https://www.ibm.com/think/topics/unsupervised-learning) uses machine learning algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns in data without the need for human intervention (hence, they are “unsupervised”).

Unsupervised learning models are used for three main tasks: clustering, association and dimensionality reduction:

* **Clustering** is a data mining technique for grouping unlabeled data based on their similarities or differences. For example, K-means clustering algorithms assign similar data points into groups, where the K value represents the size of the grouping and granularity. This technique is helpful for market segmentation, image compression, and so on.
* **Association** is another type of unsupervised learning method that uses different rules to find relationships between variables in a given data set. These methods are frequently used for market basket analysis and recommendation engines, along the lines of “Customers Who Bought This Item Also Bought” recommendations.
* **Dimensionality reduction** is a learning technique that is used when the number of features (or dimensions) in a given data set is too high. It reduces the number of data inputs to a manageable size while also preserving the data integrity. Often, this technique is used in the preprocessing data stage, such as when autoencoders remove noise from visual data to improve picture quality.

**LOGISTIC STATISTICS:**

Logistic regression is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. The article explores the fundamentals of logistic regression, it's types and implementations.

Logistic regression is used for binary [classification](https://www.geeksforgeeks.org/getting-started-with-classification/) where we use [sigmoid function](https://www.geeksforgeeks.org/derivative-of-the-sigmoid-function/), that takes input as independent variables and produces a probability value between 0 and 1.

* Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value.
* It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
* In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

Types of Logistic Regression:

On the basis of the categories, Logistic Regression can be classified into three types:

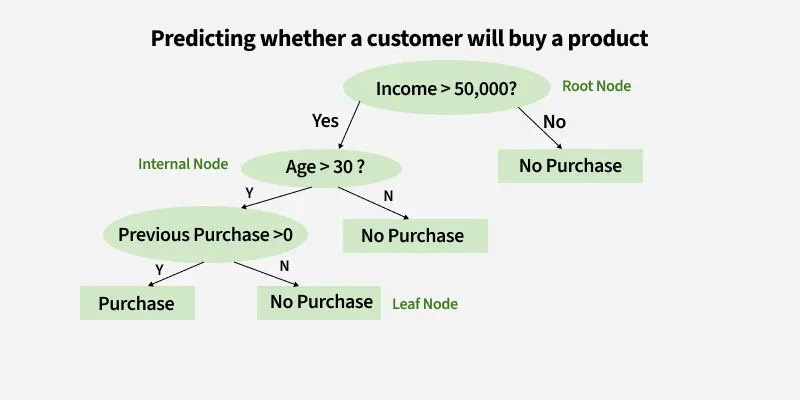
1. Binomial: In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
2. Multinomial: In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"
3. Ordinal: In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

**DECISION TREE IN MACHINE LEARNING**

A [decision tree](https://www.geeksforgeeks.org/decision-tree/) is a [supervised learning](https://www.geeksforgeeks.org/supervised-machine-learning/) algorithm used for both [classification](https://www.geeksforgeeks.org/getting-started-with-classification/) and [regression](https://www.geeksforgeeks.org/regression-in-machine-learning/) tasks. It has a hierarchical tree structure which consists of a root node, branches, internal nodes and leaf nodes. It It works like a flowchart help to make decisions step by step where:

* Internal nodes represent attribute tests
* Branches represent attribute values
* Leaf nodes represent final decisions or predictions.

Decision trees are widely used due to their interpretability, flexibility and low preprocessing needs.

****

**RANDOM FOREST:**

[Random Forest](https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/) is a method that combines the predictions of multiple decision trees to produce a more accurate and stable result. It can be used for both classification and regression tasks.

In classification tasks, Random Forest Classification predicts categorical outcomes based on the input data. It uses multiple [decision trees](https://www.geeksforgeeks.org/decision-tree/) and outputs the label that has the maximum votes among all the individual tree predictions.

Random Forest Classification works by creating multiple decision trees each trained on a random subset of data. The process begins with [Bootstrap Sampling](https://www.geeksforgeeks.org/bootstrap-method/)where random rows of data are selected with replacement to form different training datasets for each tree.

Then where only a random subset of features is used to build each tree ensuring diversity across the models.

During the training phase Feature Sampling is applied to each tree built by recursively partitioning the data based on the features. At each split the algorithm selects the best feature from the random subset optimizing for information gain or Gini impurity. The process continues until a predefined stopping criterion is met such as reaching maximum depth or having a minimum number of samples in each leaf node. After the trees are trained each tree makes a prediction. The final prediction for classification tasks is determined by majority voting.