**K-Means**

**What does it do?**

K-Means takes data points as input and groups them into k clusters. This process of grouping is the training phase of the learning algorithm. The result would be a model that takes a data sample as input and returns the cluster that the new data point belongs to, according the training that the model went through.

**Is it Supervised/Unsupervised/Reinforcement learning?**

K-means clustering is an unsupervised machine learning algorithm.

**When is it most useful?**

It has been successfully used in market segmentation, computer vision, and astronomy among many other domains.

**Random Forest**

**What does it do?**

A random forest is a data construct applied to machine learning that develops large numbers of random decision trees analysing sets of variables. This type of algorithm helps to enhance the ways that technologies analyse complex data.

The Random Forest Algorithm is composed of different decision trees, each with the same nodes, but using different data that leads to different leaves. It merges the decisions of multiple decision trees in order to find an answer, which represents the average of all these decision trees.

**Is it Supervised/Unsupervised/Reinforcement learning?**

The random forest algorithm is a supervised learning model.

**When is it most useful?**

The random forest algorithm is used in a lot of different fields, like banking, the stock market, medicine and e-commerce. In finance, for example, it is used to detect customers more likely to repay their debt on time or use a bank's services more frequently. In this domain it is also used to detect fraudsters out to scam the bank.

**Logistic Regression**

**What does it do?**

Logistic Regression is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable.

The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It’s an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

**Is it Supervised/Unsupervised/Reinforcement learning?**

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique

**When is it most useful?**

Logistic Regression is ideal for binary classification.

Examples of Logistic Regression situations include – predict whether an email is a spam, whether a tumour is malignant or not, whether it a toxic or consumable mushroom.

**Linear Regression**

**What does it do?**

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis.

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Linear Regression tries to find out the best linear relationship that describes the data you have. It assumes that there exists a linear relationship between a dependent variable and independent variable(s). The value of the dependent variable of a linear regression model is a continuous value i.e., real numbers

**Is it Supervised/Unsupervised/Reinforcement learning?**

Linear Regression is a supervised machine learning algorithm.

**When is it most useful?**

Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

**Decision Tree**

**What does it do?**

Decision trees are one of the most powerful classification algorithms that falls under supervised learning-based algorithms. It is used as a tool for making predictions and can be incorporated in different fields. With the help of decision trees, the dataset can be divided in different ways based on different conditions.

The main entities of a decision tree are the decision nodes and the leaves. The decision nodes are the ones where the data gets fragmented, whereas the leaves are one where we get the output. The concept of the decision tree can be used for both regressions as well as the classification model.

**Is it Supervised/Unsupervised/Reinforcement learning?**

A decision tree is a supervised learning technique.

**When is it most useful?**

The decision tree algorithm tries to solve a problem by using tree representation. For example, when provided with a dataset about user experience, a classification algorithm will generate a set of rules/ questions it can use to predict whether the user will convert.

**SVM (Support Vector Machine)**

**What does it do?**

A support vector machine (SVM) is machine learning algorithm that analyses data for classification and regression analysis. SVM is a method that looks at data and sorts it into one of two categories. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible.

**Is it Supervised/Unsupervised/Reinforcement learning?**

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data for classification and regression analysis.

**When is it most useful?**

Support Vector Machine algorithm is mainly used to solve classification problems. Support vectors are nothing but the coordinates of each data item. Support Vector Machine is a frontier that differentiates two classes using hyper-plane.

This algorithm is not effective for large sets of data.

**Naive Bayes**

**What does it do?**

A naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects. Naive Bayes classifiers assume strong, or naive, independence between attributes of data points. These classifiers are widely used for machine learning because they are simple to implement. Naive Bayes is also known as simple Bayes or independence Bayes.

**Is it Supervised/Unsupervised/Reinforcement learning?**

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems

**When is it most useful?**

It is mainly used in text classification that includes a high-dimensional training dataset. Popular uses of naive Bayes classifiers include spam filters, text analysis and medical diagnosis.

**KNN (K-Nearest Neighbours)**

**What does it do?**

KNN: K Nearest Neighbors is one of the fundamental algorithms in machine learning. Machine learning models use a set of input values to predict output values. KNN is one of the simplest forms of machine learning algorithms mostly used for classification. It classifies the data point on how its neighbour is classified.

KNN classifies the new data points based on the similarity measure of the earlier stored data points. For example, if we have a dataset of tomatoes and bananas. KNN will store similar measures like shape and colour. When a new object comes it will check its similarity with the colour (red or yellow) and shape.

**Is it Supervised/Unsupervised/Reinforcement learning?**

The K-Nearest Neighbors (KNN) algorithm is a simple, supervised machine learning algorithm.

**When is it most useful?**

KNN is mainly used for classification predictive problems in industry.