**Prediction Algorithms:**

It is a Supervised Machine Learning Algorithm.

It is used to predict another numerical data with the numerical data in its data.

**Classification Algorithms:**

It is a Supervised Machine Learning Algorithm.

It is used to estimate a categorical data with the numerical data in its data.

**Clustering Algorithms:**

It is an Unsupervised Machine Learning Algorithm.

Clustering is the grouping of data with similar characteristics in a data set.

There are many similarities within the same cluster, less similarities between clusters.

**Linear Regression:**

* Linear regression is a basic and widely used predictive analysis.
* Numerical input and output values are used.
* ( Cannot be used in categorical data. )
* In short, our aim is to show true values as a single truth.
* Regression equation with one dependent and one independent variable:
* y = ax + b
* y = dependent variable, a = coefficient, x = independent variable, c = constant.
* The amount of error is the distance between a point in the true value and a line.

**Multiple Linear Regression:**

* Multiple linear regression is the most widely used linear regression analysis.
* Multiple linear regression can work with more than one argument.
* For example: y = b0 + a x1 + bx2 + c a x3 + d
* For example, in Linear Regression, we predicted sales by months.
* In this algorithm, we can estimate the shoe size from the weight, age and height data.
* Here, weight, age, height are independent variables, that is, more than one.

**Polynomial Regression:**

* Data are sometimes non-linear.
* In this case, Polynomial Regression is used.

**Decision Tree:**

* Decision Tree method is an algorithm that can be used double-sided.
* It is one of the most popular machine learning algorithms.
* It is also frequently used in the field of data mining.
* Decision trees are generally enough to be thought of at the human level.
* It is very simple to understand the data and make some good comments and visualize it.
* Decision tree is a renewal process,
* As the name suggests, a tree structure is used.
* It starts with a single node and branches into new results, creating a tree structure.
* When the algorithm runs, the entered value moves on a certain path by looking at the nodes.
* And it gives a result, there are 3 kinds of knots.
* Chance Node: Indicated by a circle. Indicates multiple possible paths.
* Decision Node: Indicated by a rectangle. Indicates that a decision will be made.
* End Node: It is indicated with a triangle. Indicates a result.

**Random Forest:**

* The Random Forest algorithm is used double-sided like the decision tree.
* For Regression, the working logic dataset is divided into small pieces and trees are created.
* When predicting, the average of the predictions on the trees is taken.
* For example, 3 decision trees are created from a data set that we will make an age estimate.
* Suppose the predicted values in these decision trees are 15, 25, and 20.
* When the random forest algorithm will produce a prediction result, it takes the average of these three values.
* As a result, the predicted value is (15 +25 + 20) / 3 = 20.

**Support Vector Regression:**

* Support Vector Machines are often used for classification problems.
* It is one of the supervised learning methods.
* Draws a line to separate points placed on a plane.
* This is intended to be the maximum distance for the points of both classes.
* It is suitable for complex but small to medium sized datasets.

**Logistic Regression:**

* Logistic Regression is a regression method for classification.
* It is used to classify categorical or numerical data.
* It works only if the dependent variable, ie the result, can take 2 different values.
* (Yes / No, Male / Female, Fat / Thin etc.)
* It is widely used in linear classification problems.
* For this reason, it is very similar to Linear Regression.

**K-NN:**

* The K-NN algorithm is one of the simplest and most used classification algorithm.
* K-NN is a non-parametric, lazy learning algorithm.
* If we try to understand the concept of lazy, "eager learning" is rather lazy learning.
* There is no training phase. It does not learn training data but instead memorizes training data.
* When we want to make a guess, it looks for the closest neighbors in the entire data set.
* In the operation of the algorithm, a K value is determined.
* The meaning of this K value is the number of elements to look at.
* When a value comes, the nearest K elements are taken, the distance between the incoming value is calculated.
* The Euclidean function is generally used in the distance calculation process.
* Manhattan, Minkowski and Hamming functions can also be used as an alternative to the Euclidean function.
* After the distance is calculated, it is sorted and the corresponding value is assigned to the appropriate class.

**Gaussian Naive Bayes:**

* The Naive Bayes classifier is based on Bayes' theorem.
* It is a lazy learning algorithm and can also work on unstable datasets.
* The way the algorithm works calculates the probability of each state for an element.
* Categorizes the probability according to the highest value.
* With a little training data, he can do very successful jobs.
* If a value in the test set has an unobservable value in the training set,
* It gives 0 as a probability value, so it cannot predict.
* This condition is commonly known as Zero Frequency.
* Correction techniques can be used to resolve this situation.
* One of the simplest correction techniques is known as Laplace estimation.
* Examples of usage areas:
* real-time prediction, multi-class prediction, text classification, spam filtering ...

**K-Means Algorithm:**

* After the K value is determined, it randomly selects K center points in the algorithm.
* It calculates the distance between each data and randomly determined center points.
* Assigns data to a cluster based on the closest center point.
* Then a center point is chosen again for each cluster.
* Clustering is done according to the new center points.
* This situation continues until the system becomes stable.
* The random assignment of starting center points in the K-Means algorithm is a problem.
* David Arthur and Sergei Vassilvitskii find a way to solve this problem in 2007.
* They developed the K-Means ++ algorithm, which is a variation of the K-Means algorithm.

**Hierarchical Clustering:**

* As the name suggests, hierarchical clustering is a clustering algorithm.
* There are two different variations as Agglomerative and Divisive.
* In the agglomerative version, all data is first put into a cluster.
* So if there are N elements, N sets are formed.
* Later, clusters that are close to each other in distance merge to form a new cluster.
* This situation continues until the system is stable.
* Divisive is the opposite of Agglomerative.
* Initially all data are created in a single cluster.
* Later, this cluster is broken down and the clustering process is done.
* There are many ways to calculate distance in agglomerative hierarchical clustering.
* They are also used in creating dendrograms.

**Association Rule Mining:**

* ARM algorithms can work very successfully with categorical data.
* Market basket application is usually given as an example to Association Rule Mining.
* This process finds associations between products in customers' purchases.
* Thus, it analyzes the purchasing habits of the customers.
* For example, a customer who buys a coke also buys bread, or a customer who buys an egg also buys a wafer.
* These types of partnerships reveal the information about which products the customers buy together.
* Market managers determine the shelf layouts in the light of this information.
* They can increase sales rates and develop effective sales strategies.