•Linear Regression

In the most simple words, Linear Regression is a linear model e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y)

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.

It's a **supervised** Machine Learning model in which the model finds the best fit linear line between the independent and dependent variable.

When there is a single input variable (x), the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression.

You can perform linear regression in Microsoft Excel or use statistical software packages

Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions.

Examples :

- You can also use linear-regression analysis to try to predict a salesperson's total yearly sales (the dependent variable) from independent variables such as age, education and years of experience.

- Changes in pricing often impact consumer behavior — and linear regression can help you analyze how. For instance, if the price of a particular product keeps changing, you can use regression analysis to see whether consumption drops as the price increases.

- Linear regression techniques can be used to analyze risk. For example, an insurance company might have limited resources with which to investigate homeowners' insurance claims; with linear regression, the company's team can build a model for estimating claims costs.

•Logistic Regression

Logistic regression is named for the function used at the core of the method, the logistic function.

It is a predictive analysis algorithm and based on the concept of probability.

It's a **supervised** Machine Learning algorithm that helps fundamentally in binary classification.

It's used to classify elements of a set into two groups (such as true/false, yes/no, and so on) by calculating the probability of each element of the set - In multi-class classification, there are more than 2 classes for classifying data -.

It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

•Decision Tree

It's are a non-parametric **supervised** learning method used for classification and regression.

It's a model where the data is continuously split according to a certain parameter.

The tree can be explained by two entities, namely decision nodes and leaves.

The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.

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.

Decision tree is one of the predictive modelling approaches used in statistics, data mining and machine learning.

•Naive Bayes

It's a **supervised** learning algorithm, which is based on Bayes theorem and used for solving binary and multiclass classification problems.

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

This technique can be useful in evaluating many applications.

- Weather Forecasting: Based on temperature, humidity, pressure, etc., an organization can predict if it will be rainy/sunny/windy tomorrow.
- Fraud Analysis: Based on various bills submitted by an employee for reimbursement for expenditures on food, travel, etc., a business can predict the likelihood of fraud.

•KNN (K-Nearest Neighbours)

The k-nearest neighbors (KNN) algorithm is a simple supervised non-parametric algorithm that can be used to solve both classification and regression problems. However, it is more widely used in classification problems in the industry.

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

•Random Forest

Random forest is a **Supervised** Machine Learning Algorithm that is used widely in Classification and Regression problems.

The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

Put simply: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

Examples:

Banking: Random Forest is used in banking to predict the creditworthiness of a loan applicant. This helps the lending institution make a good decision on whether to give the customer the loan or not. Banks also use the random forest algorithm to detect fraudsters.

Health care : Health professionals use random forest systems to diagnose patients. Patients are diagnosed by assessing their previous medical history.

Stock market : Financial analysts use it to identify potential markets for stocks. It also enables them to identify the behavior of stocks.