Predictive Analytics

Predictive Analytics

- **Predictive analytics** is the branch of the advanced **analytics** which is used to make predictions about unknown future events

- **Predictive analytics** uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions

Predictive Analysis Process

Data

Data needed for predictive analytics is usually a mixture of historical and the real-time data.

Reporting/Analysis

Reporting includes creating, configuring, consolidating, organizing, formatting and summarizing. Detailed examination of the elements or structure of something

Monitoring

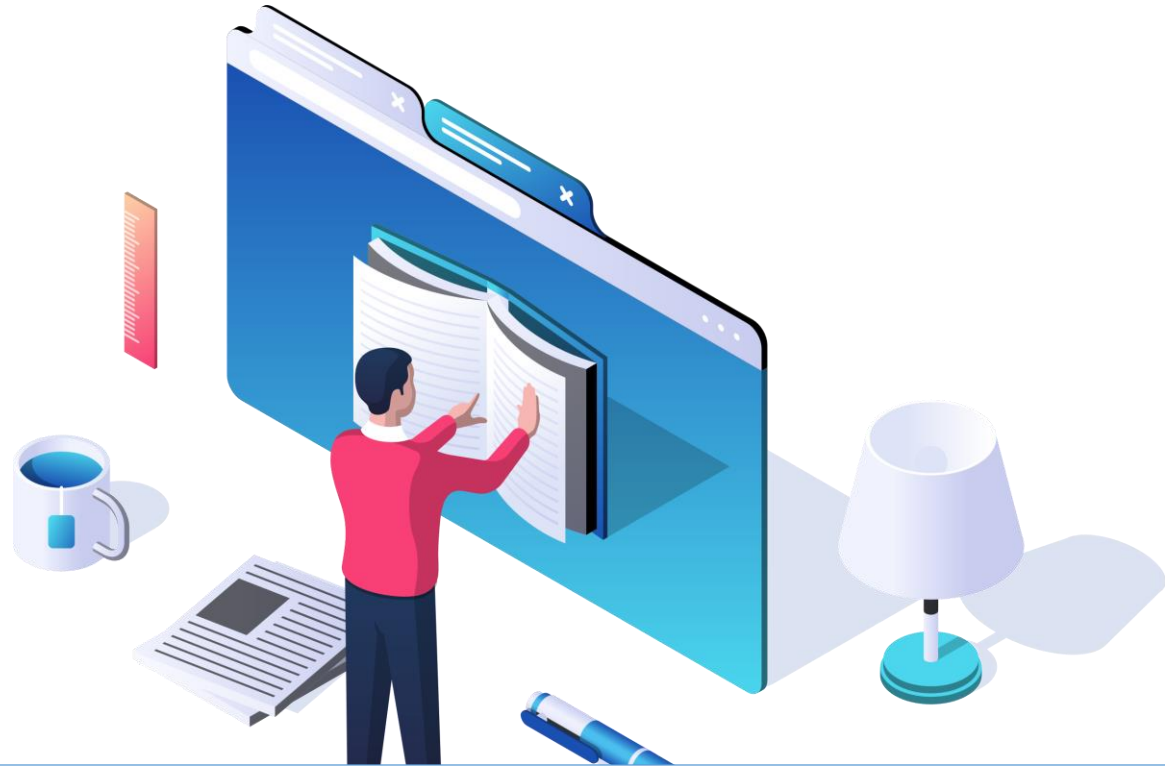
Data monitoring is a business practice in which critical business data is routinely checked against quality control rules.

Predictive Analytics

A branch of advanced analytics that makes predictions about future outcomes using historical data combined with statistical modeling, data mining techniques and machine learning.

Business process on Predictive Modeling

- Creating the model
- Testing the model
- Validating the model
- Evaluating the model



How Does The Model Works?

- In predictive modeling, data is collected for the relevant predictors, a statistical model is formulated, predictions are made and the model is validated (or revised) as additional data becomes available.
- The model may employ a simple linear equation or a complex neural network, mapped out by sophisticated software.

How Does The Model Work?

Here you will learn what a predictive model is, and how, by actively guiding marketing campaigns, it constitutes a key form of business intelligence. we'll take a look inside to see how a model works-

Predictors rank your customers to guide your marketing

Combined predictors means smarter rankings

The computer makes your model from your customer data

A simple curve shows how well your model works

Why Predictive Modelling?

- Nearly every business in competitive markets will eventually need to do predictive modeling to remain ahead of the curve.
- Predicting Modeling (also known as Predictive Analytics) is the process of automatically detecting patterns in data, then using those patterns to foretell some event.

Applications of Predictive Modelling

Analytical Customer Relationship Management

Health care

Collection Analytics

Cross-sell

Fraud Detection

Risk Management

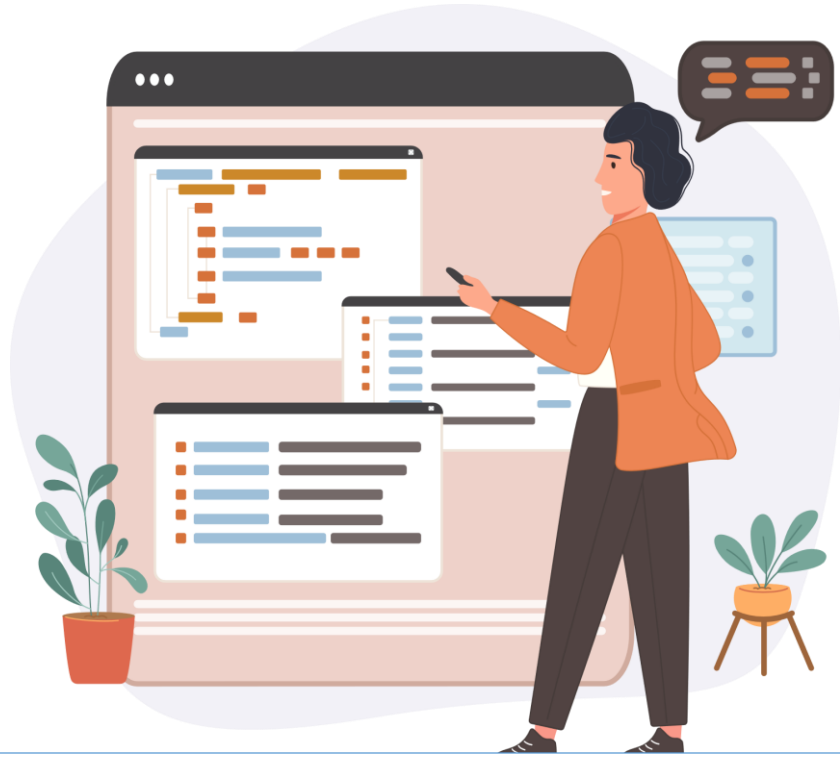
Machine Learning (Predictive Analytics)

What is Machine Learning?

Machine learning is an application of artificial intelligence that involves algorithms and data that automatically analyse and make decision by itself without human intervention.

It describes how computer perform tasks on their own by previous experiences.

Therefore, we can say in machine language artificial intelligence is generated on the basis of experience.



Types of Machine Learning

There are three types of machine learning

01

Supervised Learning

02

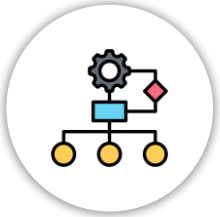
Unsupervised Learning

03

Reinforcement Learning

Classification Algorithm

Classification

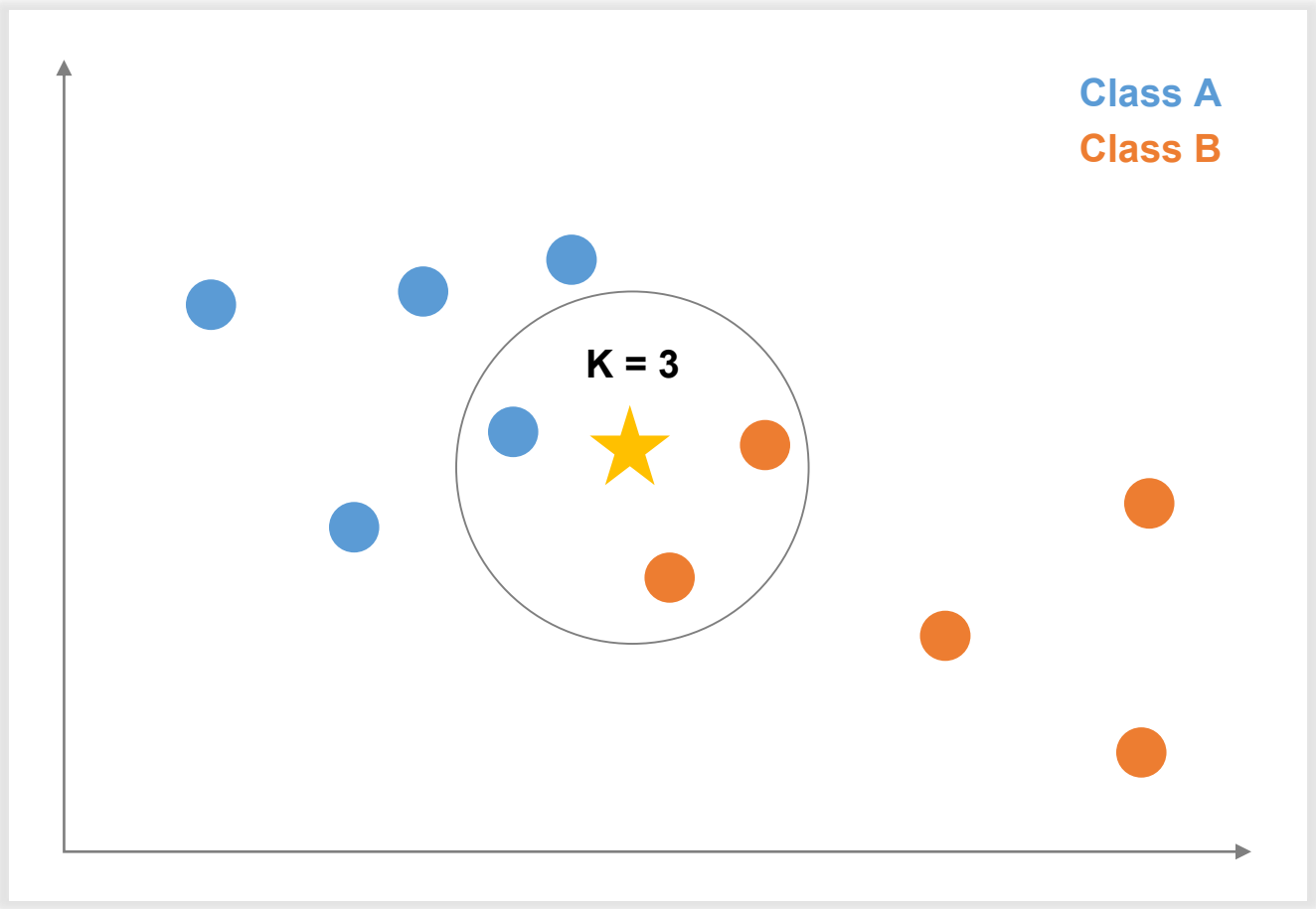


Classification algorithms used in machine learning utilize input training data for the purpose of predicting the likelihood or probability that the data that follows will fall into one of the predetermined categories.

One of the most common applications of classification is for filtering emails into “spam” or “non-spam”, as used by today’s top email service providers.

K-nearest neighbors algorithm (KNN)

- It is a supervised machine learning algorithm.
- The algorithm can be used to solve both classification and regression problem statements.
- The number of nearest neighbors to a new unknown variable that has to be predicted or classified is denoted by the symbol 'K



Decision Tree

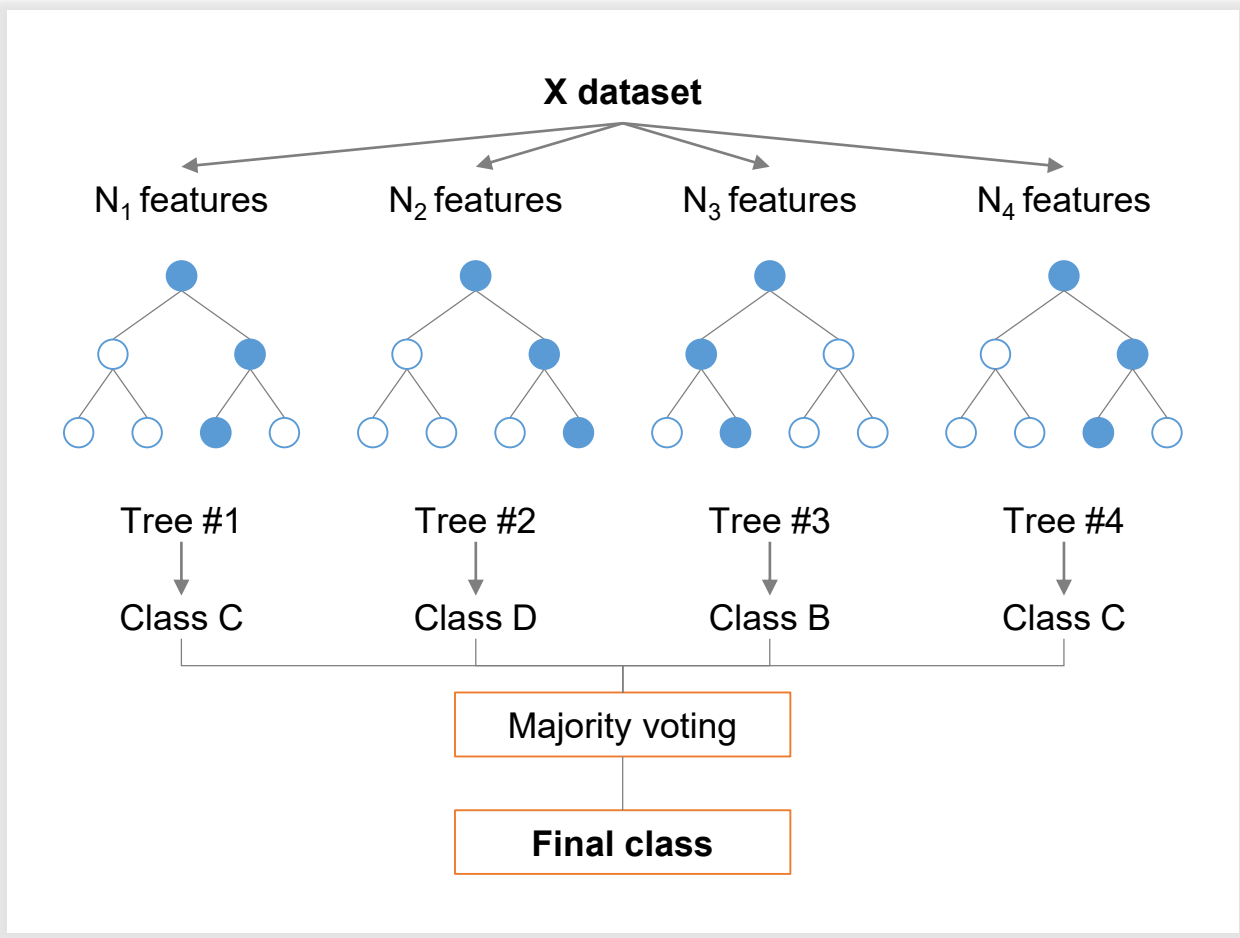
A decision tree is a very specific type of probability tree that enables you to make a decision about some kind of process.

For example, you might want to choose between manufacturing item A or item B, or investing in choice 1, choice 2, or choice 3.

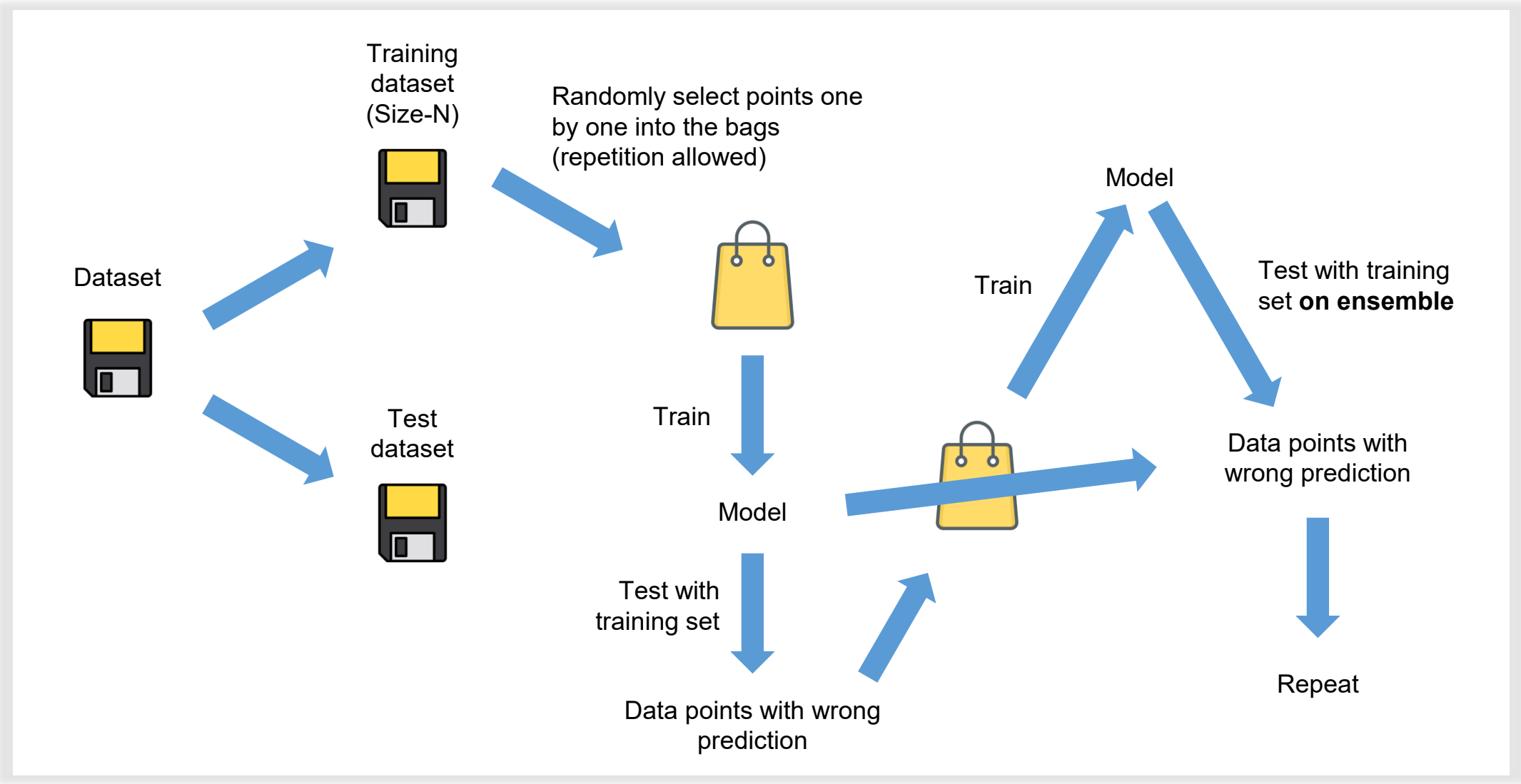
Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

For classification tasks, the output of the random forest is the class selected by most trees



Boosting


Dataset
Training dataset (Size-N)
Test dataset
Randomly select points one by one into the bags (repetition allowed)
Train
Model
Test with training set
Data points with wrong prediction
Train
Model
Test with training set **on ensemble**
Data points with wrong prediction
Repeat

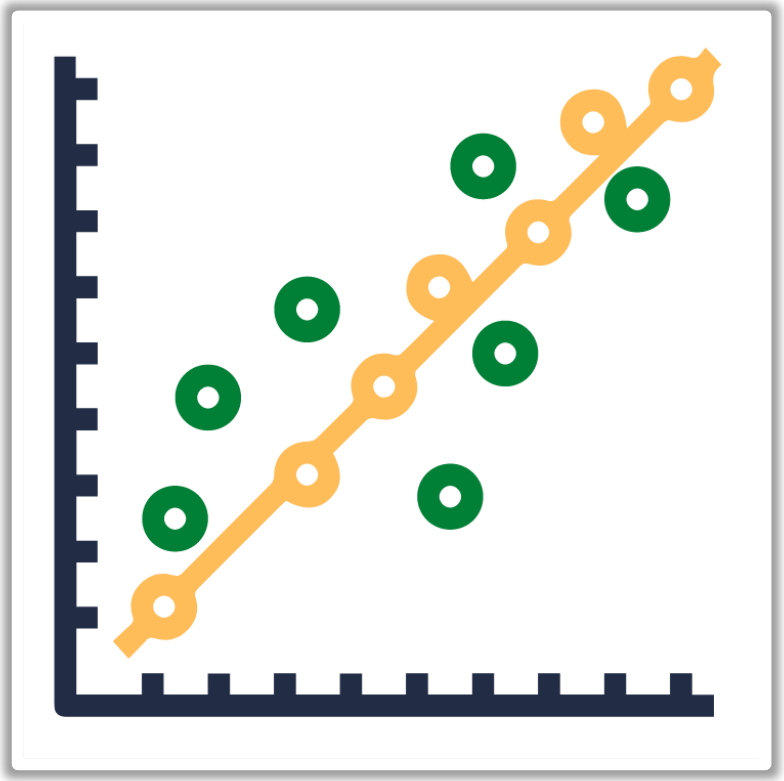
Regression Algorithm

Regression

Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables

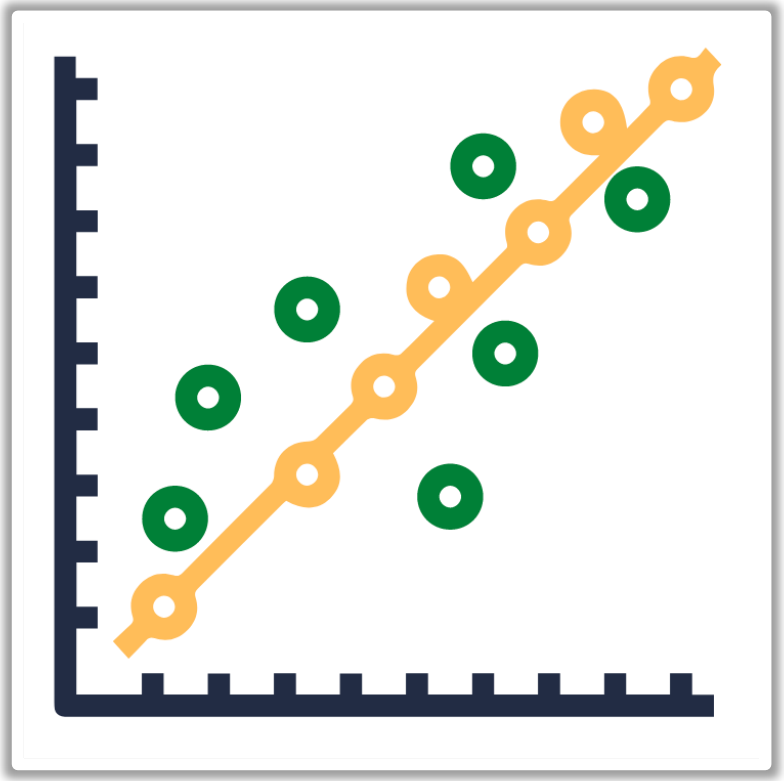
More specifically, Regression analysis helps us to understand how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are held fixed.

It predicts continuous/real values such as **temperature, age, salary, price**, etc



Types of Regression

1. Simple Linear Regression
2. Multiple Linear Regression
3. Polynomial Regression



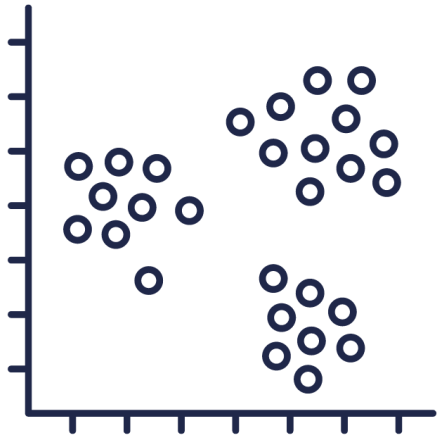
Clustering Algorithm

Clustering Algorithms

- Clustering or cluster analysis is a machine learning technique, which groups the unlabelled dataset.
- It can be defined as "A way of grouping the data points into different clusters, consisting of similar data points.
- The objects with the possible similarities remain in a group that has less or no similarities with another group."

The clustering technique can be widely used in various tasks. Some most common uses of this technique are

- Market Segmentation
- Statistical data analysis
- Social network analysis
- Image segmentation
- Anomaly detection, etc.



Time Series Forecasting

Time Series Forecasting

- A **time series** is a series of data points indexed (or listed or graphed) in time order.
- Most commonly, a time series is a sequence taken at successive equally spaced points in time.
- Thus, it is a sequence of discrete-time data.
- Examples of time series are heights of ocean tides, counts of sunspots, and the daily closing value of the Dow Jones Industrial Average.

Thank you!

Any questions?