**Machine Learning**

It is a science of making computers learn and act like humans by feeding data and information without being explicitly programmed. While writing a piece of code in a program, we give instructions to a computer, and it follows them step by step. However, in machine learning, the system learns on its own; we only provide past information.

# Process:

* Data gathering
* Data preprocessing
* Choose a model
* Train model (iterative)
* Test model (iterative)
* Tune Model

# Real world Application:

* Face recognition
* Health Industry
* Weather forecasting
* Producing a web series

# Types:

## Supervised learning:

The machine is trained on a set of labeled data, which means that the input data is paired with the desired output. The machine then learns to predict the output for new input data. Supervised learning is often used for tasks such as classification, regression, and object detection.

|  |  |
| --- | --- |
| **Classification** | **Regression** |
| Fraud detection | Weather forecasting |
| Email spam detection | Risk assessment |
| Image classification | Score prediction |

## Unsupervised learning:

The machine is trained on a set of unlabeled data, which means that the input data is not paired with the desired output. The machine then learns to find patterns and relationships in the data. Unsupervised learning is often used for tasks such as [clustering](https://www.geeksforgeeks.org/clustering-in-machine-learning/), dimensionality reduction, and anomaly detection.

|  |  |
| --- | --- |
| **Association** | **Clustering** |
| Market basket analysis | City planning |
| Text mining | Target marketing |
| Face recognition | Medical research |

## Reinforcement learning:

 Reinforcement learning differs from supervised learning in a way that in supervised learning the training data has the answer key with it so the model is trained with the correct answer itself whereas in reinforcement learning, there is no answer but the reinforcement agent decides what to do to perform the given task. In the absence of a training dataset, it is bound to learn from its experience.

|  |
| --- |
| Gaming |
| Robot Navigation |
| Stock trading |
| Assembly line progresses |

# Algorithms:

Regression is a statistical method used to model and analyze the relationships between a dependent variable and one or more independent variables. The primary goal of regression analysis is to understand the nature of the relationship and make predictions.

## Linear regression:

Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. The goal is to find the best-fitting straight line through the data points that minimizes the sum of the squared differences between the observed values and the values predicted by the line.



## Logistic regression:

Logistic regression is a statistical method used for modeling the probability of a binary (or dichotomous) outcome based on one or more predictor variables. Unlike linear regression, which predicts a continuous outcome, logistic regression predicts the probability of a categorical outcome, usually coded as 0 or 1.

## Decision tree:

A decision tree is a popular machine learning algorithm used for both classification and regression tasks. It models decisions and their possible consequences as a tree-like structure of nodes, branches, and leaves. Decision trees are intuitive, easy to interpret, and can handle both numerical and categorical data.

## Random forest:

Random Forest is an ensemble learning method in data science that is used for both classification and regression tasks. It builds multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random Forests are popular due to their high accuracy, ability to handle a large number of features, and robustness to overfitting.

## K Nearest Neighbors:

K-Nearest Neighbors (KNN) is a simple, non-parametric, and intuitive algorithm used for both classification and regression tasks in data science. It relies on the concept of proximity, where predictions are made based on the closest training examples in the feature space.