# EXPERIMENT – 01

## WRITE 6 MACHINE LEARNING ALGORITHMS WITH THERE DESCRIPTION AND ONE APPLICATION OF EACH.

- ❖ **Linear Regression:**

  - ➕ **Description:**
    - Linear Regression Is A Supervised Learning Algorithm and Tries to Model the Relationship Between A Continuous Target Variable and One Or More Independent Variables by Fitting A Linear Equation to The Data.
    - A Linear Regression Model Tries to Fit A Regression Line to The Data Points That Best Represents the Relations or Correlations.
    - Equation:

    $$y=A+Bx$$

    - Where Y Is the Dependent Variable, x Is the Independent Variable, B Is the Slope of The Line and A Is The y-Intercept.

  - ➕ **Example:**
    - Suppose We Would Like to Design A Computing Machinery That Can Estimate the Price of Used Vehicles, Say A Car.
    - The Feature Such as The Brand, Year, Capacity, Etc., Are Taken as Input and The Output Is the Price of The Cars.
    - In These Problem Regression Can Help Use in Estimating the Car Price.

- ❖ **K-Means Clustering:**

  - ➕ **Description:**
    - Clustering Is A Way to Group A Set of Data Points in A Way That Similar Data Points Are Grouped Together.
    - K-Means Clustering Aims to Partition Data into K Clusters in A Way That Data Points in The Same Cluster Are Similar and Data Points in The Different Clusters Are Farther Apart.
    - Firstly, We Have to Randomly Select Centroids for Each Cluster. Calculate the Distance of All Data Points to The Centroids.
    - Assign Data Points to The Closest Cluster. Find the New Centroids of Each Cluster by Taking the Mean of All Data Points in The Cluster.
    - Repeat Steps 2,3 And 4 Until All Points Converge and Cluster Centres Stop Moving.
  - ➕ **Example:**
    - Clustering Helps Marketers Improve Their Customer Base and Work on The Target Areas. It Helps Group People (According to Different Criteria's Such as Willingness, Purchasing Power Etc.)

*Based on Their Similarity in Many Ways Related to The Product Under Consideration.*

- *Clustering Helps in Identification of Groups of Houses on The Basis of Their Value, Type and Geographical Locations.*
- *Clustering Is Used to Study Earth-Quake. Based on The Areas Hit by An Earthquake in A Region, Clustering Can Help Analyse the Next Probable Location Where Earthquake Can Occur.*

## ❖ Naïve Bayes:

### ➕ Description:

- *Naive Bayes Is A Supervised Learning Algorithm Used for Classification Tasks. Hence, It Is Also Called Naive Bayes Classifier.*
- *Naive Bayes Assumes That Features Are Independent Of Each Other And There Is No Correlation Between Features.*
- *The Intuition Behind Naive Bayes Algorithm Is the Bayes' Theorem:*

**P(A|B) =P(A). P(B) / P(B) (Bayes' Theorem)**

- *P(A|B): Probability of Event A Given Event B Has Already Occurred*
- *P(B|A): Probability of Event B Given Event A Has Already Occurred*
- *P(A): Probability of Event A*
- *P(B): Probability of Event B*

### ➕ Example:

- *For Example, The Temperature Being 'Hot' Has Nothing to Do with The Humidity or The Outlook Being 'Rainy' Has No Effect on The Winds.*
- *Hence, The Features Are Assumed to Be Independent.*

## ❖ Logistic Regression:

### ➕ Description:

- *Logistic Regression Is A Supervised Learning Algorithm Which Is Mostly Used for Binary Classification Problems.*
- *The Basis of Logistic Regression Is the Logistic Function, Also Called the Sigmoid Function, Which Takes in Any Real Valued Number and Maps It to A Value Between 0 And 1.*
- *Equation:*

**1/ (1 + E^-Value)**

### ➕ Example:

- *For Example, You May Want to Know the Likelihood of a Visitor Choosing an Offer Made on Your Website — Or Not (Dependent Variable).*
- *Your Analysis Can Look at Known Characteristics of Visitors, Such as Sites They Came From, Repeat Visits to Your Site, Behaviour on Your Site (Independent Variables).*
- *Logistic Regression Models Help You Determine A Probability of What Type of Visitors Are Likely to Accept the Offer — Or Not.*

- As A Result, You Can Make Better Decisions About Promoting Your Offer or Make Decisions About the Offer Itself.

## ❖ K-Nearest Neighbours:

### ♦ Description:
- K-Nearest Neighbours (KNN) Is A **Supervised** Learning Algorithm That Can Be Used to Solve Both Classification and Regression Tasks.
- The Main Idea Behind KNN Is That the Value or Class of a Data Point Is Determined by The Data Points Around It.
- KNN Classifier Determines the Class of a Data Point by Majority Voting Principle.
- It Is Very Important to Determine an Optimal K Value. If K Is Too Low, The Model Is Too Specific and Not Generalized Well. It Also Tends to Be Sensitive to Noise. This Situation Is Known as Overfitting.
- On the Other Hand, If K Is Too Large, The Model Is Too Generalized and Not A Good Predictor on Both Train and Test Sets. This Situation Is Known as Underfitting.

### ♦ Example:
- KNN Can Be Effectively Used in Detecting Outliers.
- One Such Example Is Credit Card Fraud Detection.

## ❖ Decision Tree:

### ♦ Description:
- A Decision Tree Builds Upon Iteratively Asking Questions to Partition Data.
- In This the Data Is Visualized in The Form of Decision Tree
- The Aim of The Decision Tree Algorithm Is to Increase the Predictiveness As Much as Possible So That the Model Keeps Gaining Information About The Dataset.

### ♦ Example: An Example of The Medical Use of Decision Trees Is in The Diagnosis of a Medical Condition from The Pattern of Symptoms, In Which the Classes Defined by The Decision Tree Could Either Be Different Clinical Subtypes or A Condition, Or Patients with A Condition Who Should Receive Different Therapies.