**Machine Learning:**

It is a subset of Artificial intelligence that allows the machine to automatically learn to form past data & experience while identifying patterns to make predictions with minimal human intervention.

**Types of Machine Learning:**

1. Supervised learning
2. Unsupervised learning
3. Semi-supervised Machine Learning
4. Reinforcement Learning

**Supervised Learning:**

* When the model gets trained in “labeled data” labeled data has both input and output parameters.
* In this algorithm learn to map points between input and correct output. It has both training and validation data sets.

**Definition:**

It uses a labeled data set to train the algorithm to predict the outcomes of the target variable.

**Algorithms in supervised Learning:**

1. Linear regression
2. Logistic Regression
3. Decision Tree
4. Random Forest
5. KNN
6. Naïve Bayes

**Linear Regression:**

Error = actual – predicted

**Mean square error (MSE)**

**Root mean square error (RMSE)**

**Mean absolute percentile**

**Mean absolute percentile error**

**Regression:** This problem involve predicting a continuous, numerical values based on input data

Ex- prediction of sales, price, no. of unit

**Logistic Regression:**

When we have classification problem.

Y(target variable): class or boolboolean values.

Classification: this problem involve prediction of categorical label or class based on input data.

Why logistic Regression not Classification?

**Probability(Logistic regression)** = exp(y)/1+exp(y)

**True positive** = Machine(1 || 0) + Pred(0 || 1)

**True Negative** = Machine(1 || 0) + Pred(1 || 0)

**Accuracy** = (TP+TN)/No of instances (values)

**Recall** = TP/(TP+FN)

**Precision** is the ratio of true pos to the total pred pos.

It measures how many of the predicted values

**Decision Tree:**

It is a tool used to make decisions. It looks like a flowchart where each branch represents a choice and each choice and each leaf represent an outcome.

**Entropy**:- -p1\*Log(p1, Base2) - p2\*Log(p2, Base2)…….

**Hyperparameter tuning**

It involves selecting the best set of hyperparameters to improve the model's performance on a given data

**Parameters:**

1. **Min sample leaf:**

It is the minimum no of nodes should be present at the leaf. Higher values can prevent the model from overfitting.

1. **Min sample split:**  
   The minimum no of samples is required to split internal nodes. Higher values can prevent the model from learning.
2. **Max depth:**

The maximum of the tree limiting the depth can prevent

Overfitting.

**Random Forest:**

1. It is an ensemble learning method primarily used for classification and regression tasks.
2. It operates by constructing multiple decision trees during the training and testing model of the individual tree's **classification** and **mean prediction** (regression).

**Naïve Bayes:**

It is used for classification tasks such as text classification it uses principles of probability to perform classification tasks.

**Bernoulli NB** –

It is assumed that all our features are binary, taking only two values 1 and 0.

**Multinomial NB –**

It is used when we have discrete data ex: ratings or classes.

**Gaussian NB –**

Because of the assumption of the normal distribution this algo is used in cases when all features are in continuous form.

Explore all three Methods online\*\*