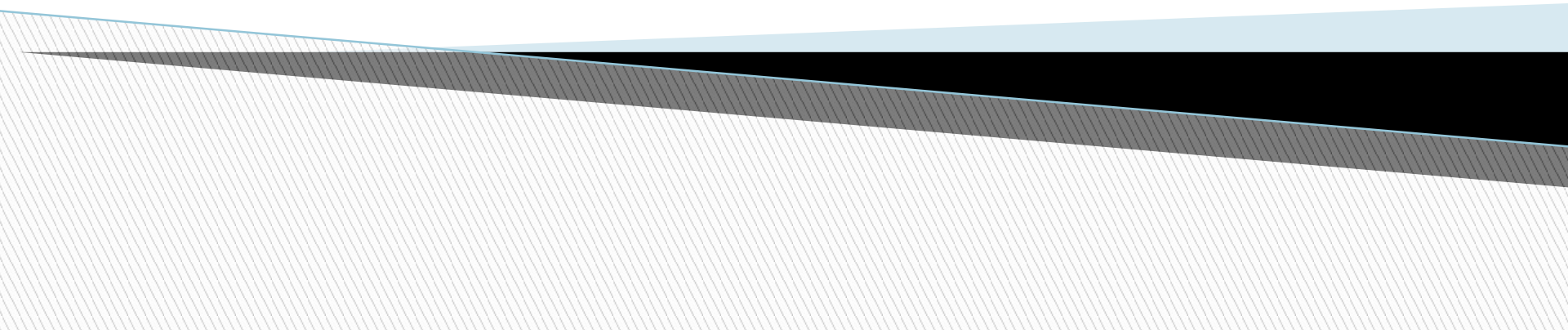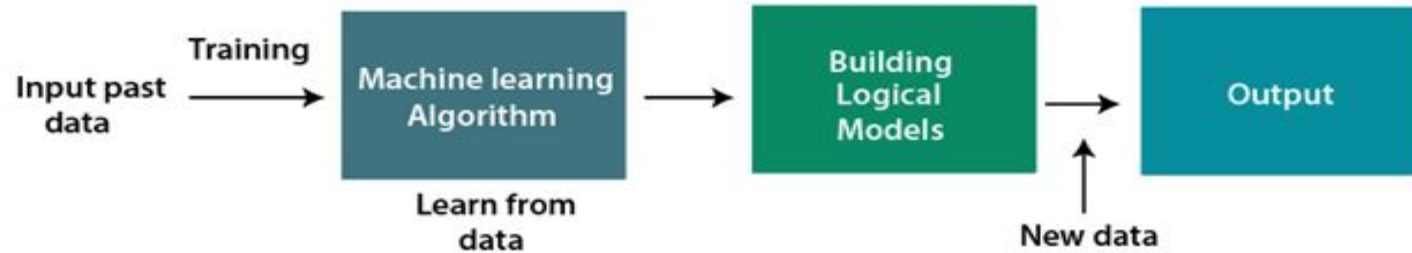# What is machine learning?

- Machine learning is a branch of AI which enables machines to learn from past data or experiences without being explicitly programmed.

- Machine learning enables a computer system to make predictions or take some decisions using historical data without being explicitly programmed.

.

Machine Learning Process

- A Machine Learning system **learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it**..

Goal of machine Learning

- The primary objective of machine learning research is to develop general-purpose algorithms in practical value.
- The main purpose of machine learning is to study and design the algorithms that can be used to produce the predicates from the given dataset.

Machine learning can also be used to:
Compete Intelligently
Enhance Customer Service
Manage our Sales Funnel
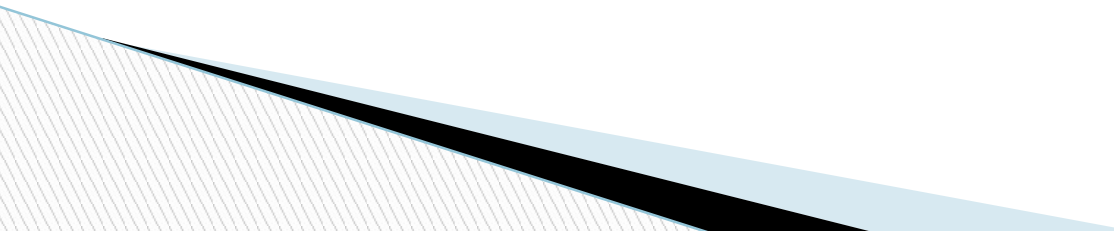Detect Fraudulent Activity
Predict Journey times
Predict how long jobs may take
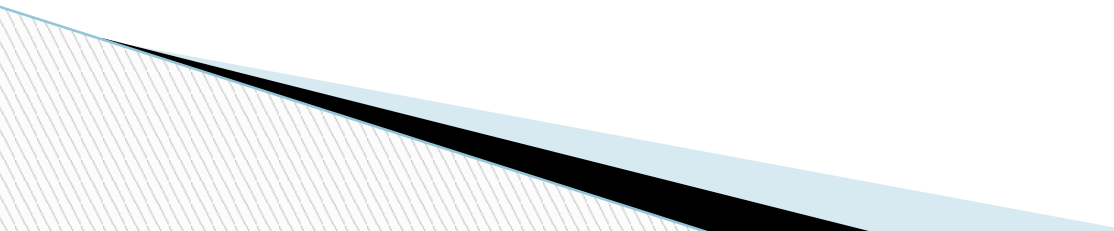Score our prospects and customers
Behavior of our market

Applications of Machine learning

1. Image Recognition
2. Speech Recognition
3. Traffic prediction
4. Product recommendations
5. Self-driving cars
6. Email Spam and Malware Filtering
7. Virtual Personal Assistant
8. Online Fraud Detection
9. Stock Market trading
10. Medical Diagnosis

Challenges for Machine Learning

1.Data Collection

2. Less Amount of Training Data

3. Non-representative Training Data

4. Poor Quality of Data

5. Irrelevant/Unwanted Features

Classification of Machine Learning
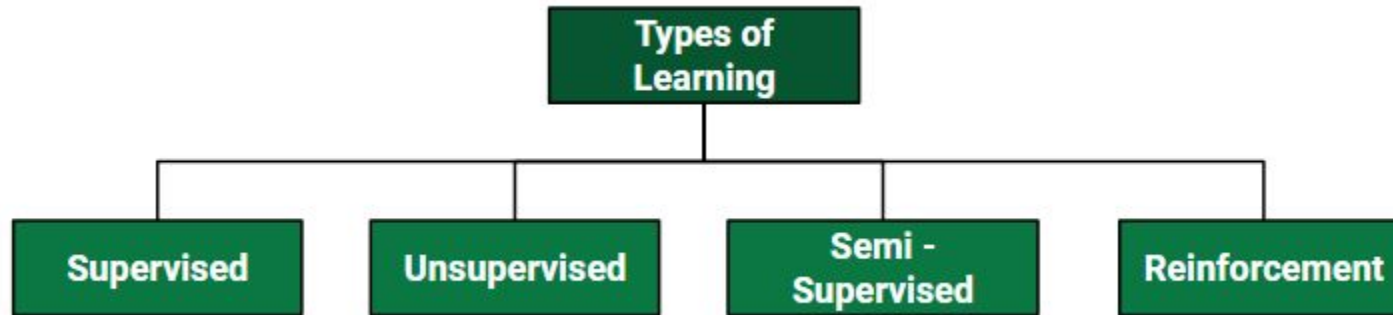
1) Supervised Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.

2) Unsupervised Learning

Unsupervised learning is a learning method in which a machine learns without any supervision.

3. Semi-supervised Learning:

its working lies between Supervised and Unsupervised techniques. We use these techniques when we are dealing with a data which is a little bit labeled and rest large portion of it is unlabeled. This technique is mostly applicable in case of image data-sets where usually all images are not labeled.

4) Reinforcement Learning

It is a feedback-based learning method. in which a learning agent gets a reward for each right action and gets a penalty for each wrong action


REINFORCEMENT

1.Supervised Learning

In Supervised Learning, a machine is trained using **'labeled' data**. Datasets are said to be labeled when they contain both input and output parameters. and on that basis, it predicts the output.

This implies that some data is already tagged with the correct answer.

The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher.
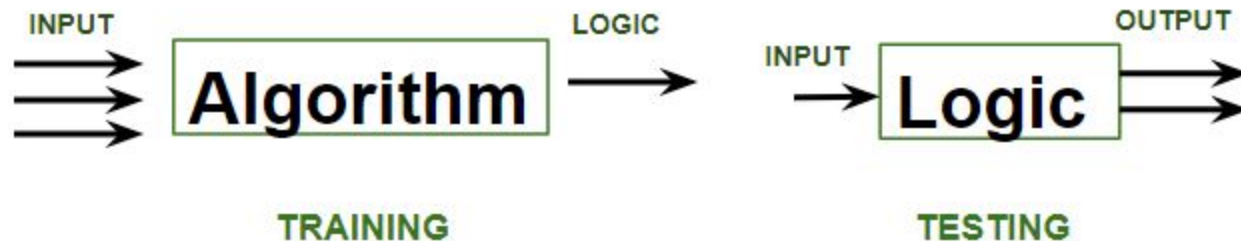
☐ Example: a basket filled with different kinds of fruits. Now the first step is to train the machine with all different fruits one by one like this:

If the shape of the object is rounded and has a depression at the top, is red in color, then it will be labeled as –**Apple**.
If the shape of the object is a long curving cylinder having Green-Yellow color, then it will be labeled as –**Banana**.

Thus the machine learns the things from training data(basket containing fruits) and then applies the knowledge to test data(new fruit).
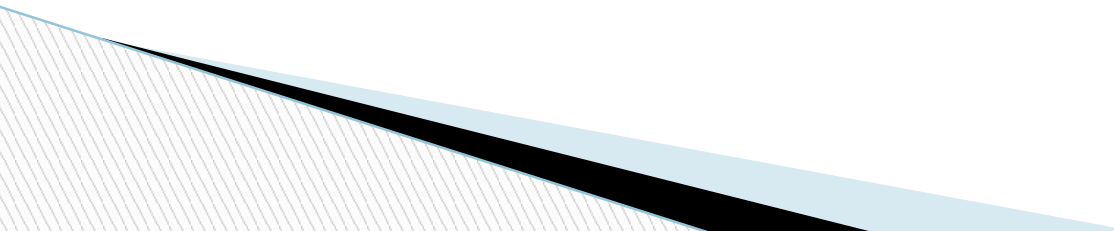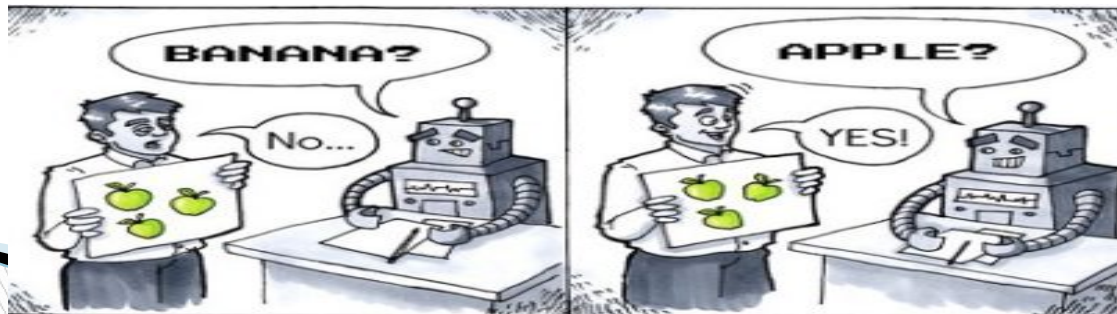
Supervised learning classified into two categories of algorithms:

1.Classification: A classification problem is when the output variable is a category, such as "Red" or "blue" or "disease" and "no disease".

2.Regression: A regression problem is when the output variable is a real value, such as "dollars" or "weight".

☐ Supervised learning classified into following categories of algorithms:

Types:-

1. Regression
2. Logistic Regression
3. Classification
4. Naive Bayes Classifiers
5. K-NN (k nearest neighbors)
6. Decision Trees
7. Support Vector Machine

2.Unsupervised learning

- Unsupervised learning is a learning method in which a machine learns without any supervision.
- The training is provided to the machine with the set of data that has not been labeled, classified.
- **No teacher** is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself.
- Unsupervised learning is classified into two categories of algorithms:

  1.Clustering: A clustering problem is where we want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

2.Association: An association rule learning problem is where we want to discover rules that describe large portions of our data, such as people that buy X also tend to buy Y.

Example: suppose it is given an image having both dogs and cats which it has never seen.
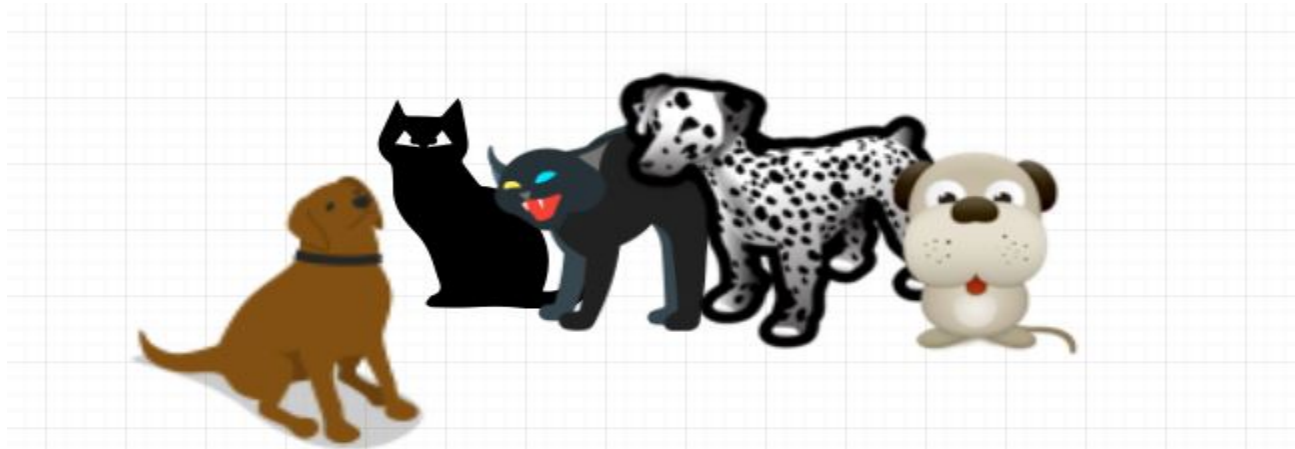
So the machine has no idea about the features of dogs and cats so we can't categorize it as 'dogs and cats '. But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts. The first may contain all pics having **dogs** in it and the second part may contain all pics having **cats** in it. Here machine didn't learn anything before, which means no training data .

Types of Unsupervised Learning

1. **Clustering**
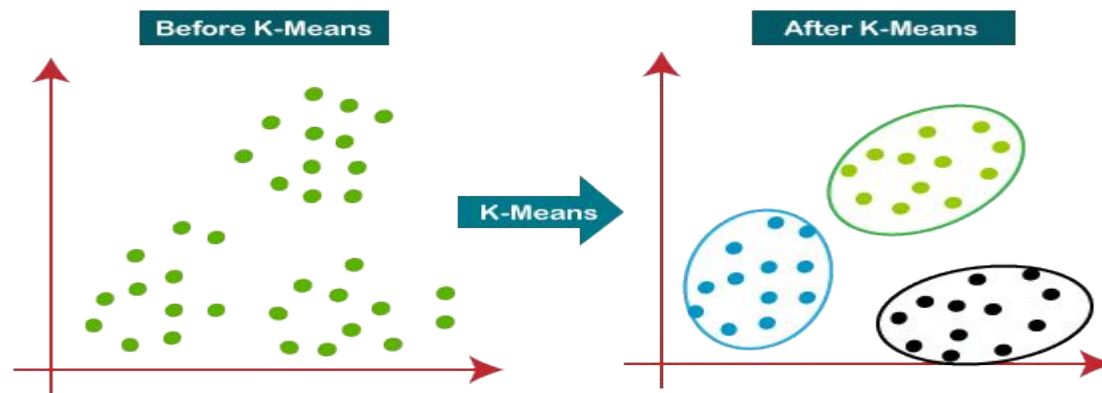2. **Association**

1.Clustering :-

- In this Learning we find patterns in the data that we are working on.

- It may be the shape, size, color etc. which can be used to group data items or create clusters.
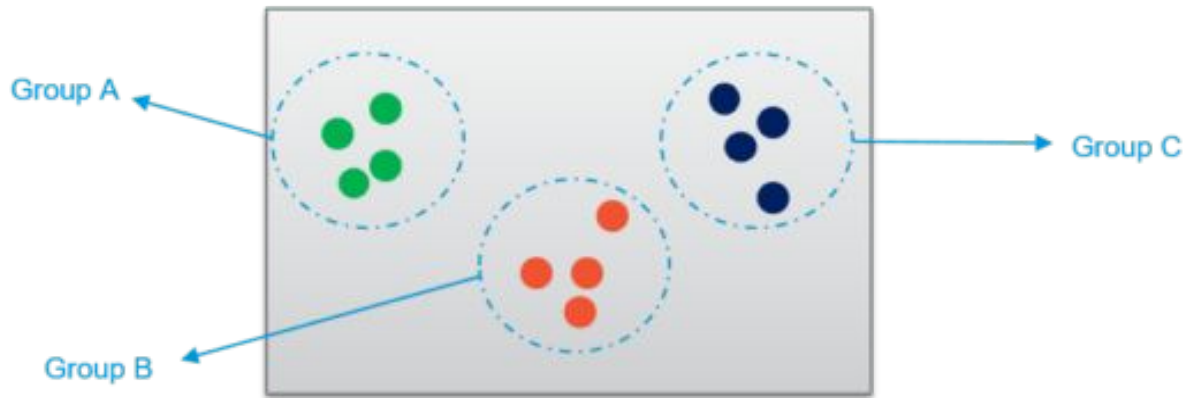
Some popular algorithms in Clustering are

1.Hierarchical Clustering – This algorithm builds clusters based on the similarity between different data points in the dataset.

2.K-Means Clustering –This algorithm groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

3.K-NN Clustering– This algorithm is also called as a lazy learner because it learns only when the algorithm is given a new data point. It works well with smaller datasets as huge datasets take time to learn.

2. **Association** :- In this type of Learning we find the dependencies of one data item to another data item and map them such that they help you profit better.

A **IF** ⟹ B **THEN**

1. **Apriori algorithm–**

- It is based on **breadth-first search**.

- This Algorithm maps the dependency of one data item with another which can help us understand what data item influences the possibility of something happening to the other data item.

- For example, bread influences the buyer to buy milk and eggs. So that mapping helps increase profits for the store.

☐ **2.FP-Growth Algorithm –**

• The Frequency Pattern (FP) algorithm finds the count of the pattern that has been repeated, adds that to a table and then finds the most possible item and sets that as the root of the tree.

• This algorithm is faster than Apriori as the support is calculated and checked for increasing iterations rather than creating a rule and checking the support from the dataset.

Supervised vs. Unsupervised Learning

Parameter	Supervised Learning	Unsupervised Learning
Dataset	Labelled	Unlabelled
Method of Learning	Guided learning	The algorithm learns by itself using dataset
Complexity	Simpler method	Computationally complex
Accuracy	More Accurate	Less Accurate