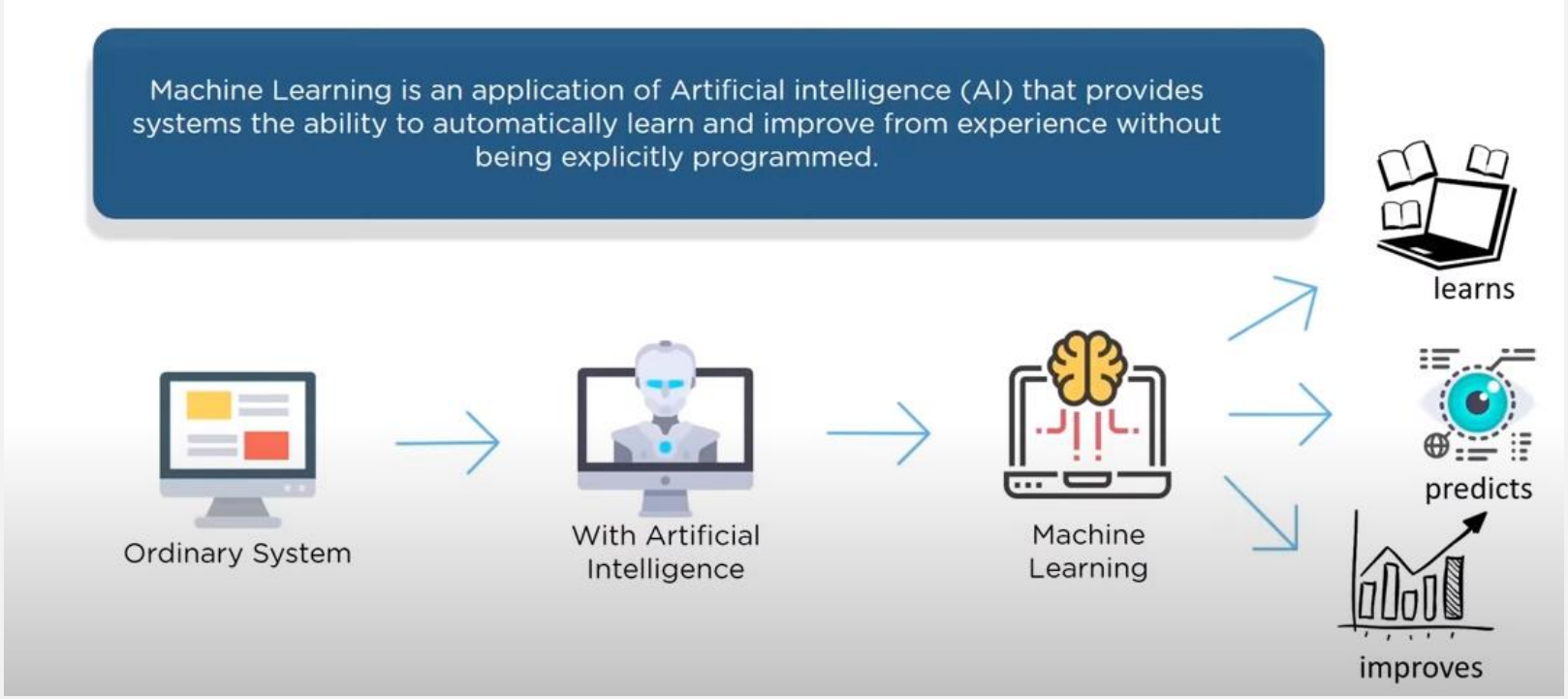
Python for Beginners

Archer Infotech , PUNE

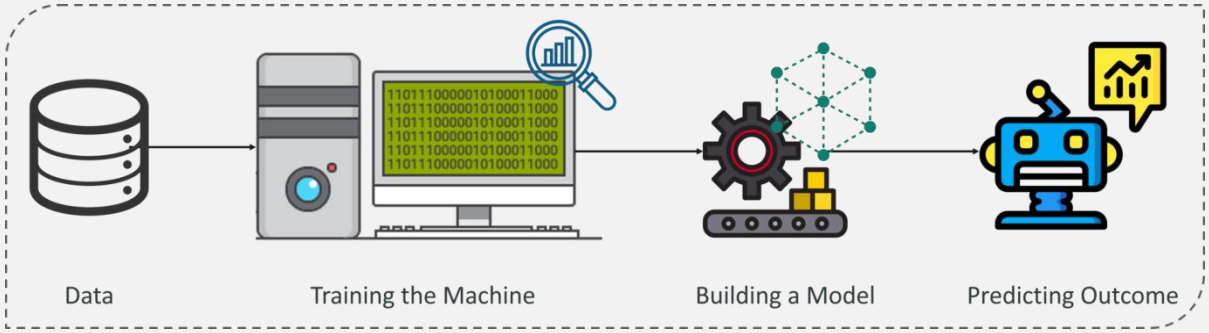
Python – Machine Learning

What is Machine Learning ?

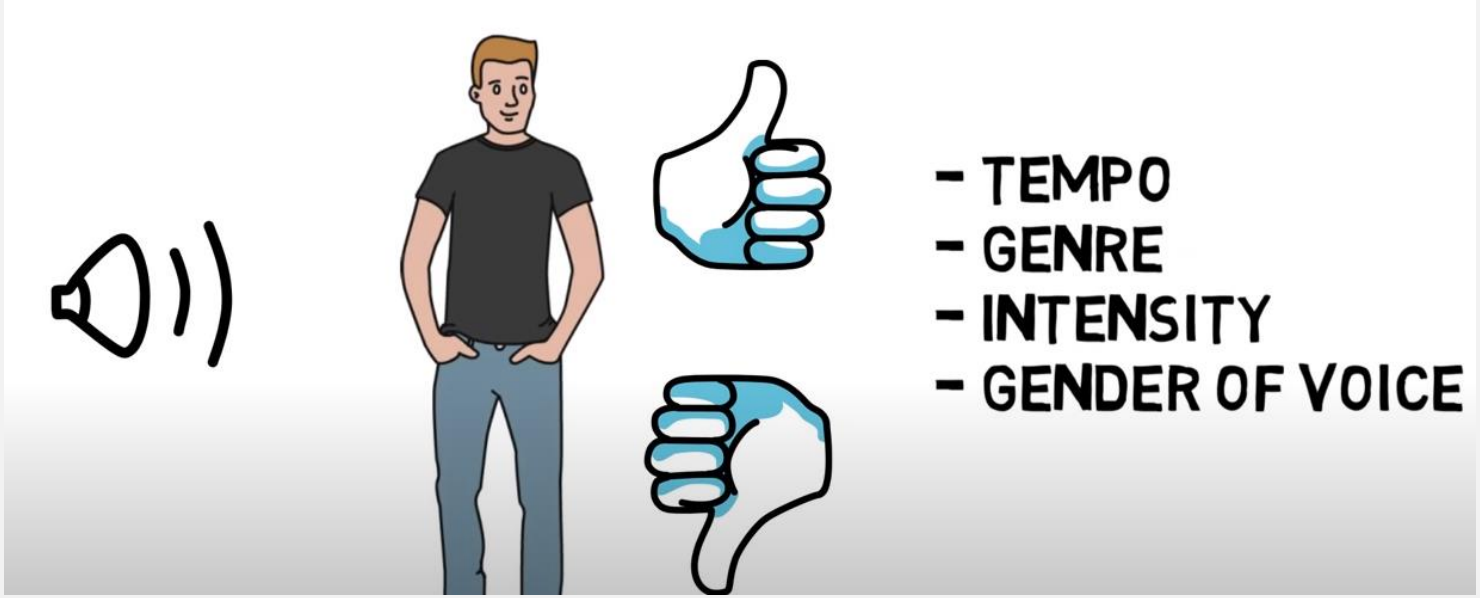


What is Machine Learning ?

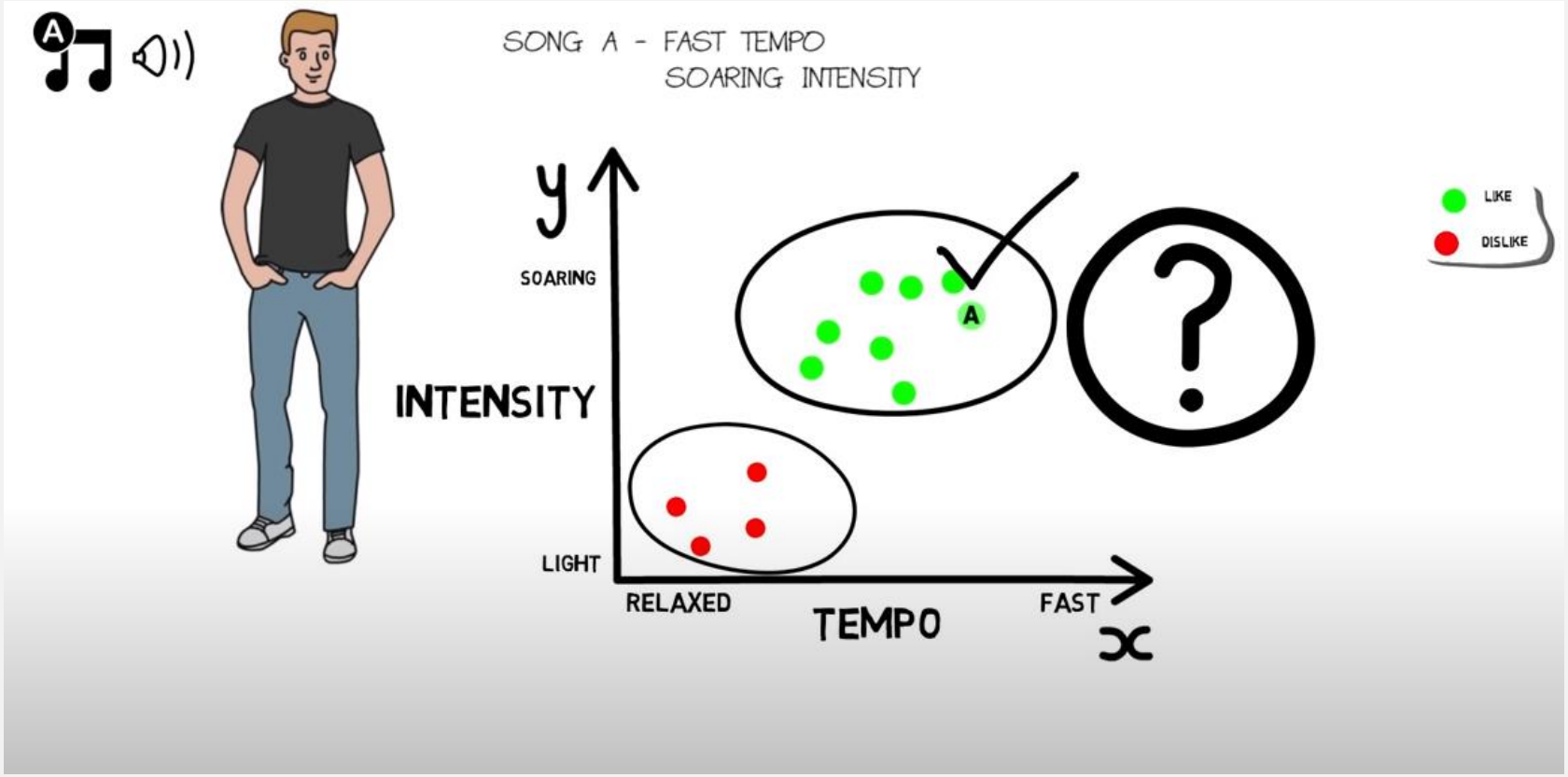
- ***Machine learning is a subset of Artificial Intelligence (AI) which provides machines the ability to learn automatically & improve from experience without being explicitly programmed to do so.*** **In the sense, it is the practice of getting Machines to solve problems by gaining the ability to think.**



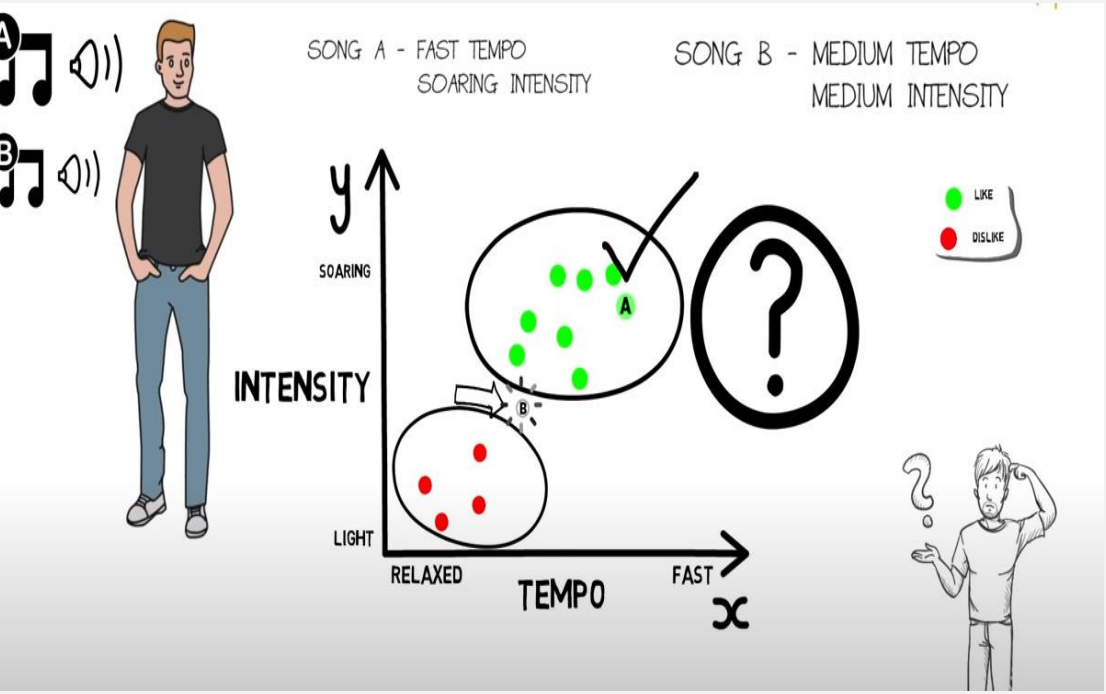
Machine Learning Example

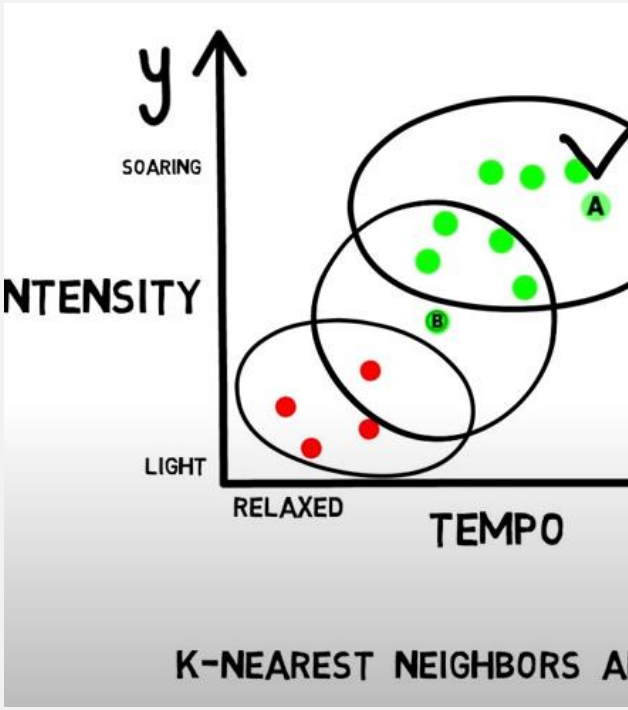


Machine Learning Example



Machine Learning Example





Machine Learning Terms

- **Algorithm: A Machine Learning algorithm is a set of rules and statistical techniques used to learn patterns from data and draw significant information from it. It is the logic behind a Machine Learning model. An example of a Machine Learning algorithm is the Linear Regression algorithm.**

- **Model: A model is the main component of Machine Learning. A model is trained by using a Machine Learning Algorithm. An algorithm maps all the decisions that a model is supposed to take based on the given input, in order to get the correct output.**

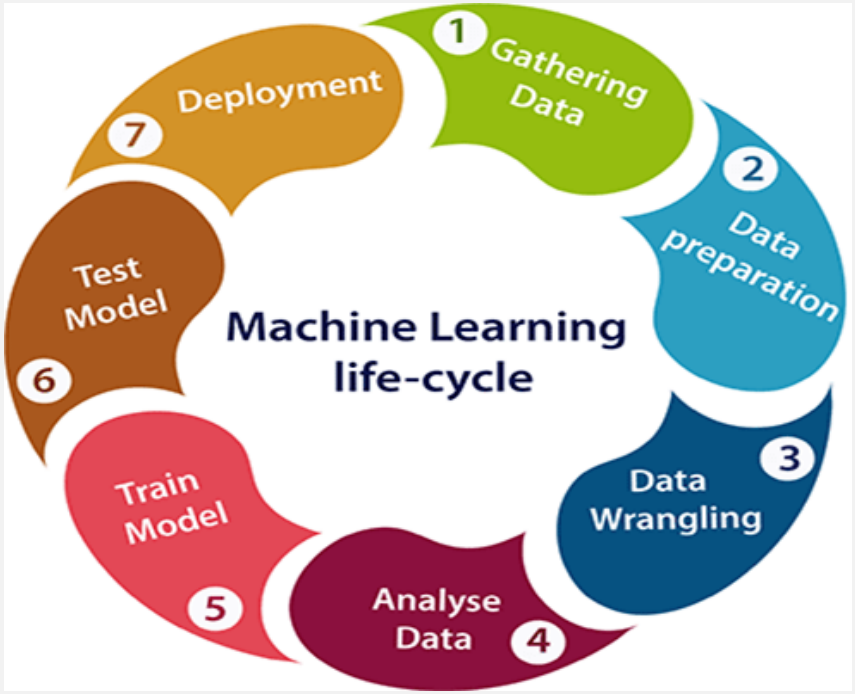
Machine Learning Terms

- **Predictor Variable:** It is a feature(s) of the data that can be used to predict the output.

- **Response Variable:** It is the feature or the output variable that needs to be predicted by using the predictor variable(s)
- .
- **Training Data:** The Machine Learning model is built using the training data. The training data helps the model to identify key trends and patterns essential to predict the output.

- **Testing Data:** After the model is trained, it must be tested to evaluate how accurately it can predict an outcome. This is done by the testing data set.

Machine Learning Process



Step 1: Define the objective of the Problem Statement
Step 2: Data Gathering
Step 3: Data Preparation
Step 4: Exploratory Data Analysis
Step 5: Building a Machine Learning Model
Step 6: Model Evaluation & Optimization
Step 7: Predictions

Machine Learning Process – Gathering Data

This step includes the below tasks:

Identify various data sources
Collect data
Integrate the data obtained from different sources

Popular sources for Machine Learning datasets

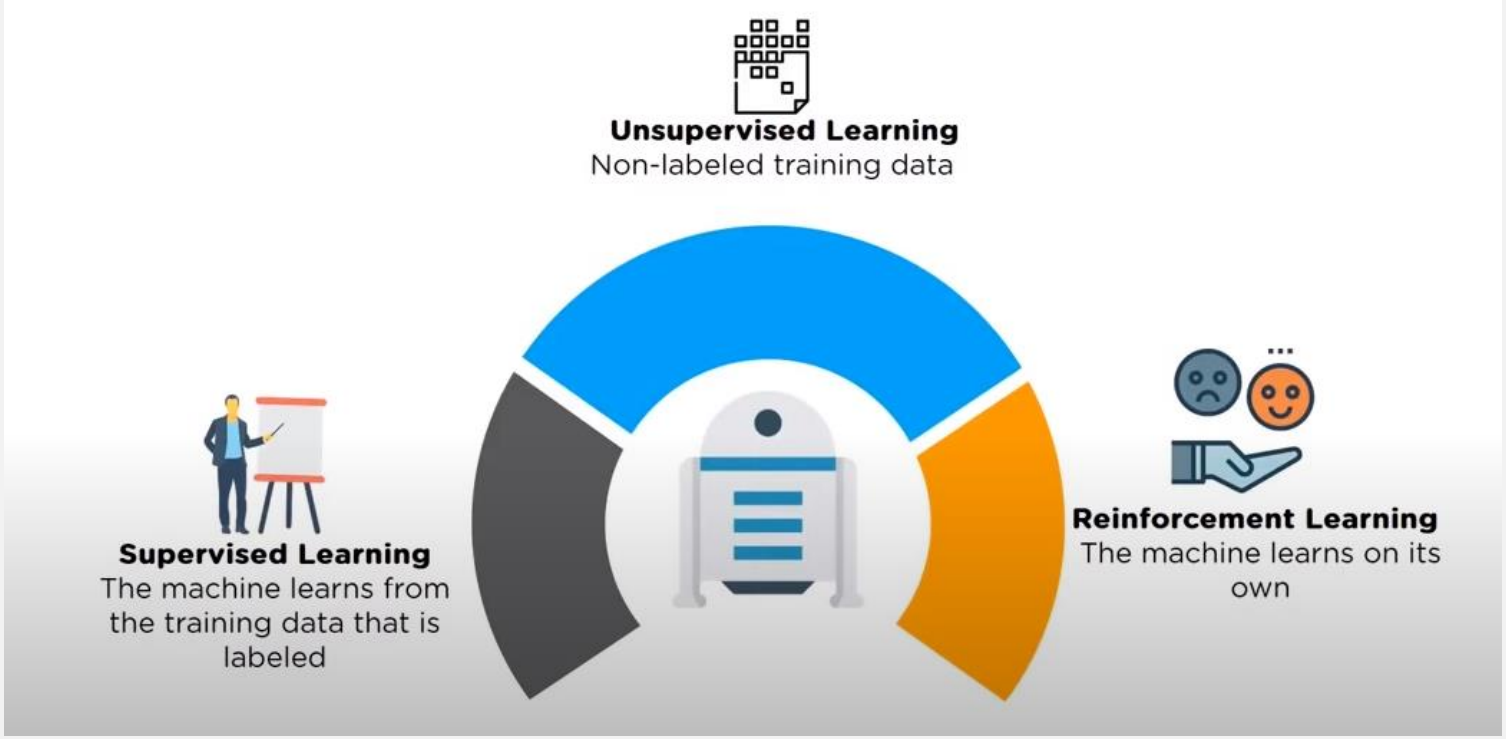
1. Kaggle Datasets
2. UCI Machine Learning Repository
3. Datasets via AWS
4. Google's Dataset Search Engine
5. Microsoft Datasets

Machine Learning Process – Data Preprocessing

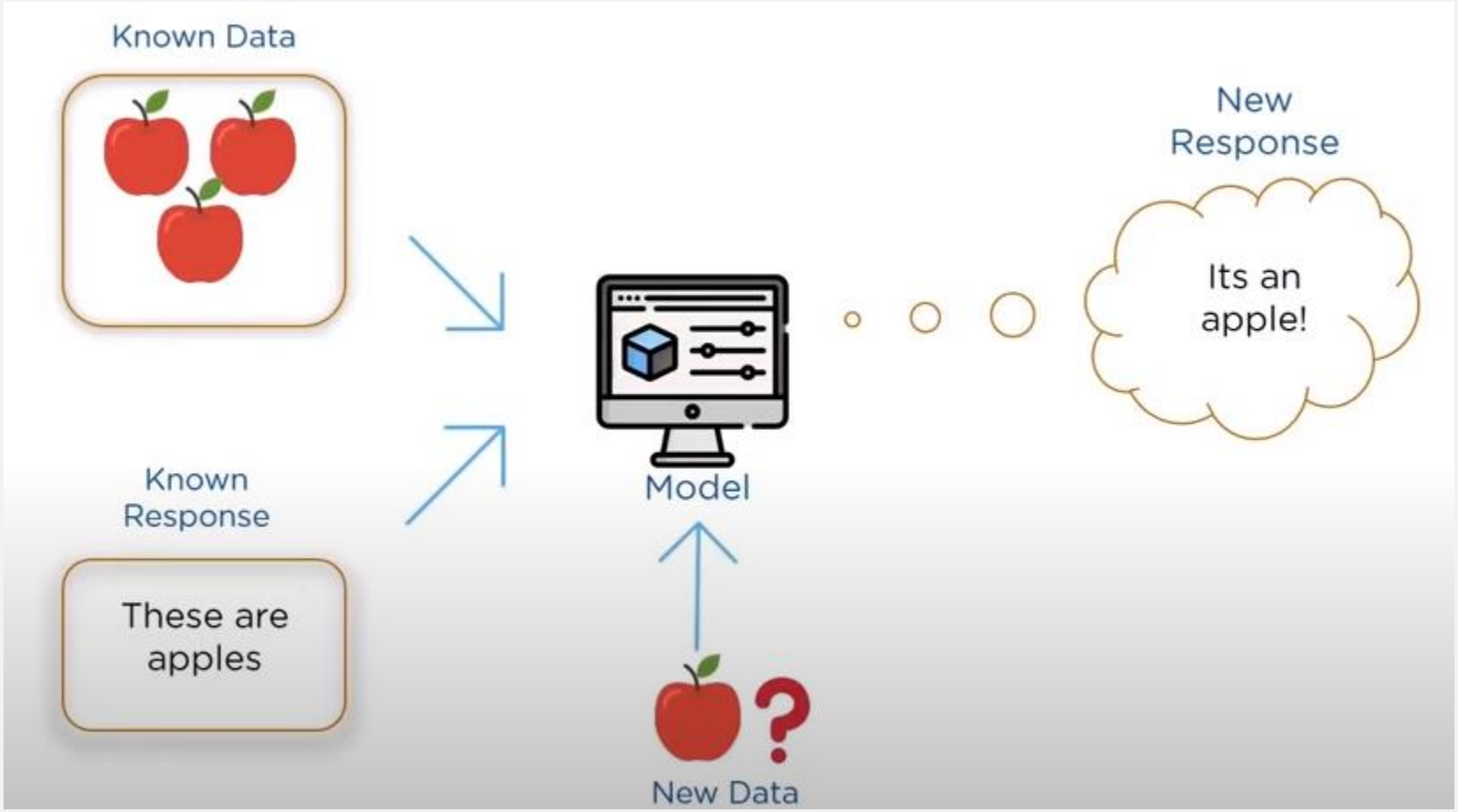
This step includes the below tasks:

Getting the dataset
Importing libraries
Importing datasets
Finding Missing Data
Encoding Categorical Data
Splitting dataset into training and test set
Feature scaling

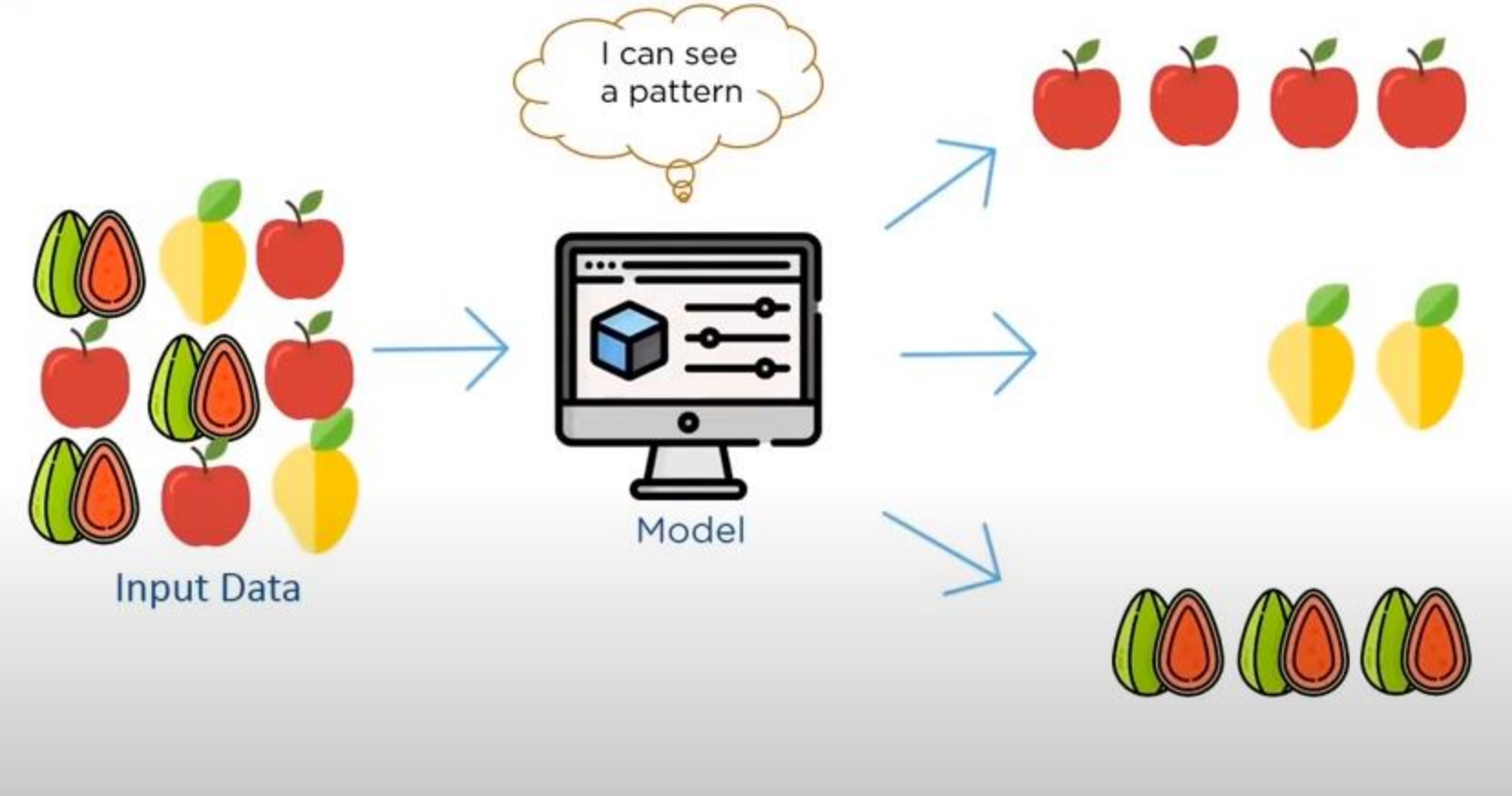
Types of Machine Learning



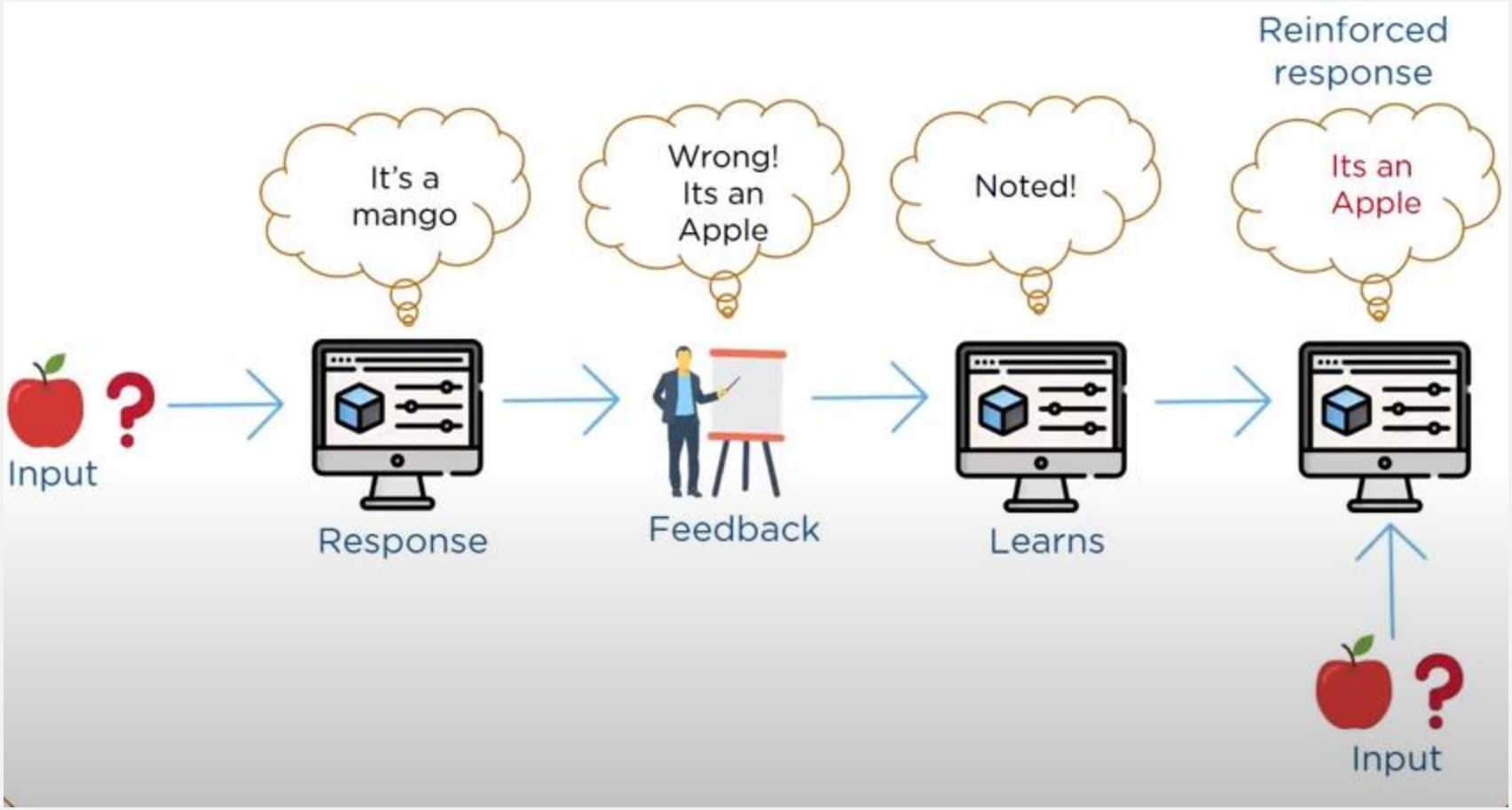
Supervised Machine Learning



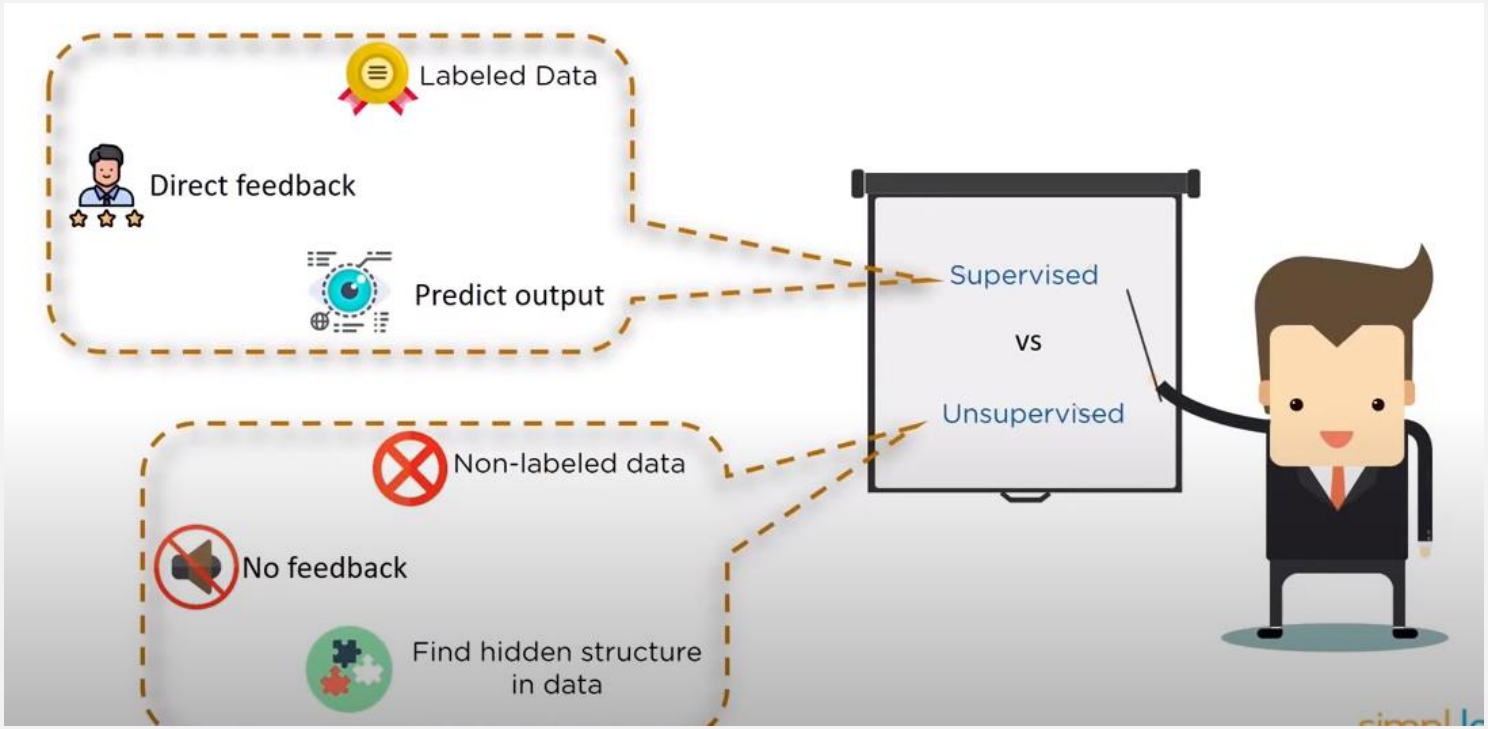
Unsupervised Machine Learning



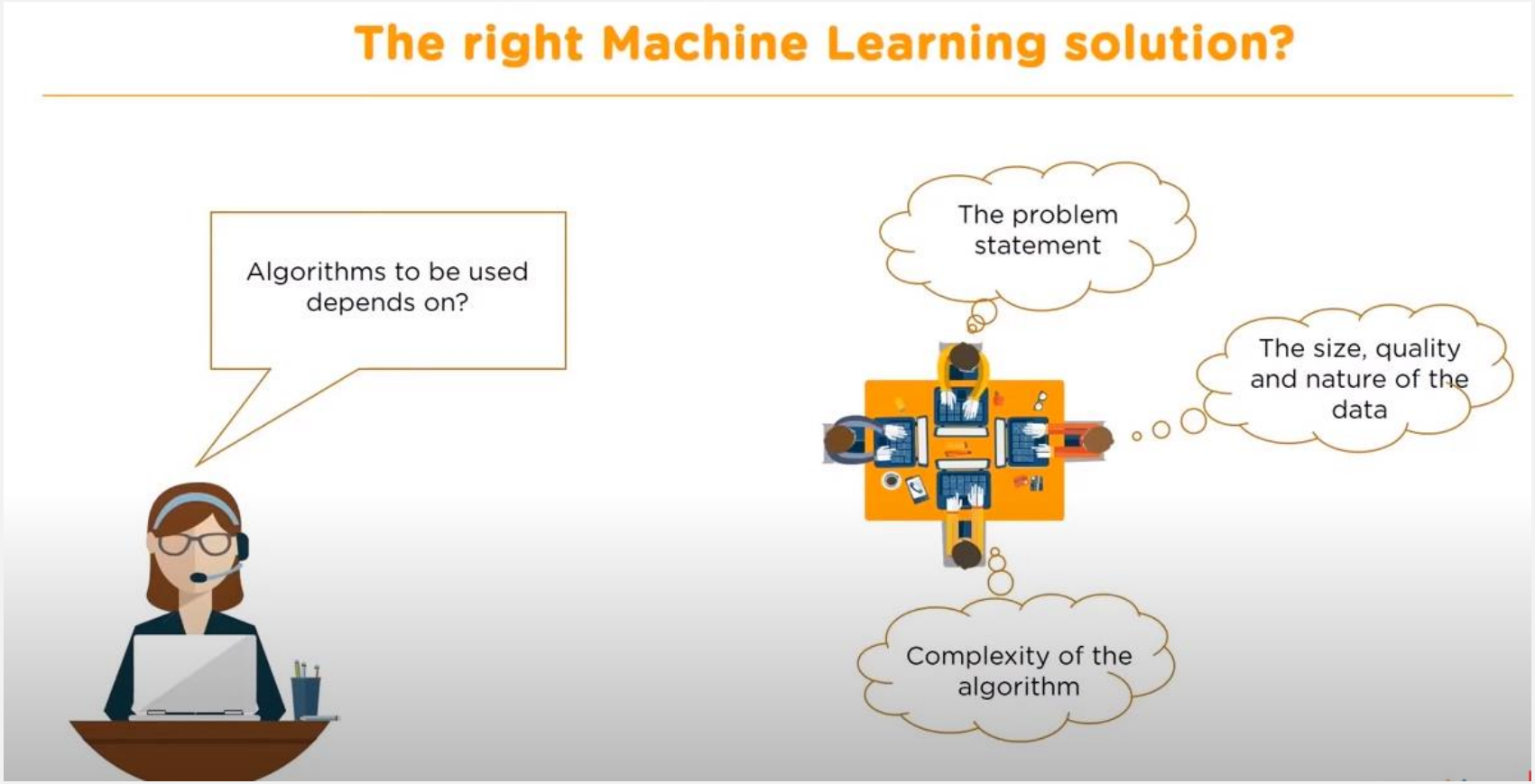
Reinforced Machine Learning



Supervised Vs. Unsupervised



Supervised Vs. Unsupervised



Supervised Vs. Unsupervised

The right Machine Learning solution?

Classification

Used when the output is categorical like 'YES' or 'NO'

Algorithms used

- Decision Tree
- Naïve Bayes
- Random Forest
- Logistic regression
- KNN

Regression

Used when a value needs to be predicted like the 'stock prices'

Algorithms used

- Linear Regression

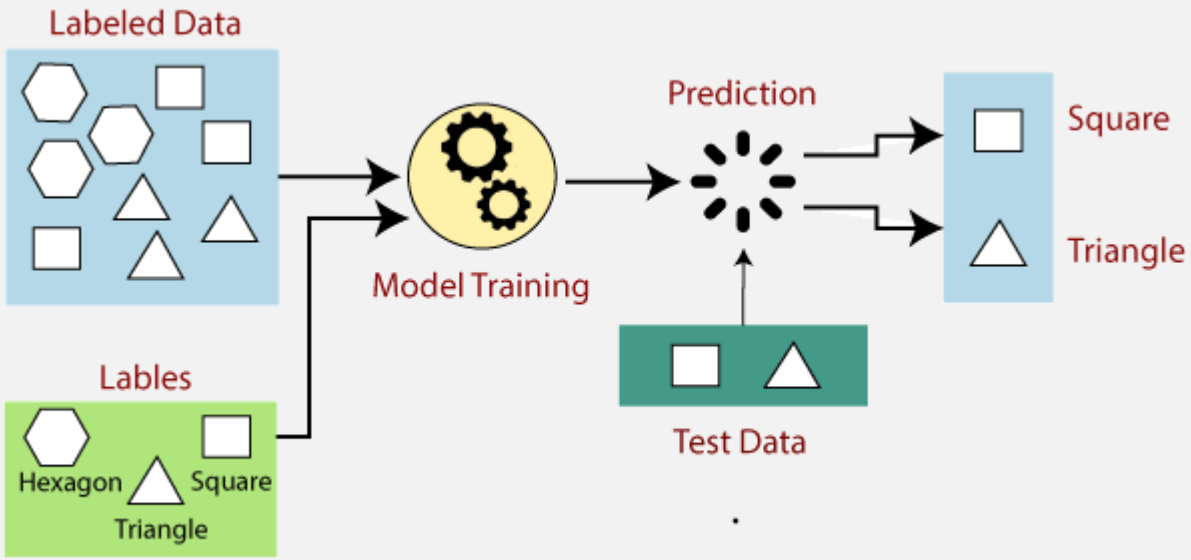
Clustering

Used when the data needs to be organized to find patterns in the case of 'product recommendation'

Algorithms used

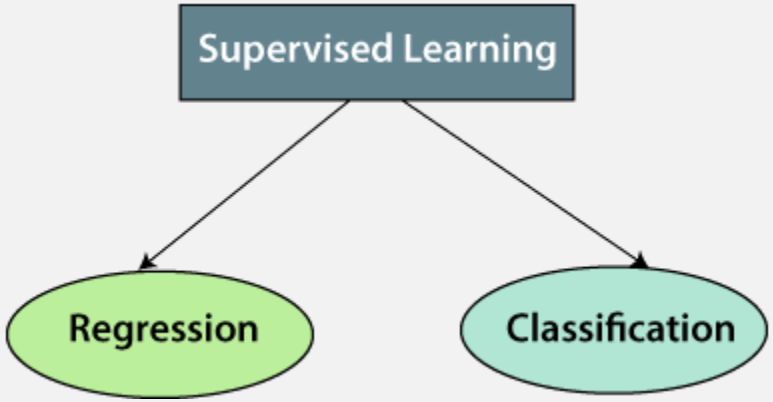
- K Means

Supervised Learning



In supervised learning, we use known or labeled data for the training data. Since the data is known, the learning is, therefore, supervised, i.e., directed into successful execution. The input data goes through the Machine Learning algorithm and is used to train the model. Once the model is trained based on the known data, you can use unknown data into the model and get a new response.

Supervised Learning



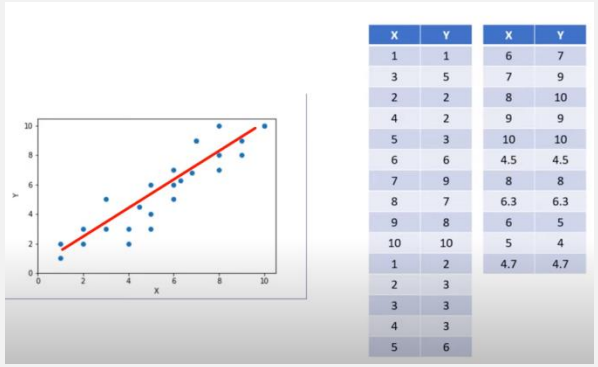
Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.

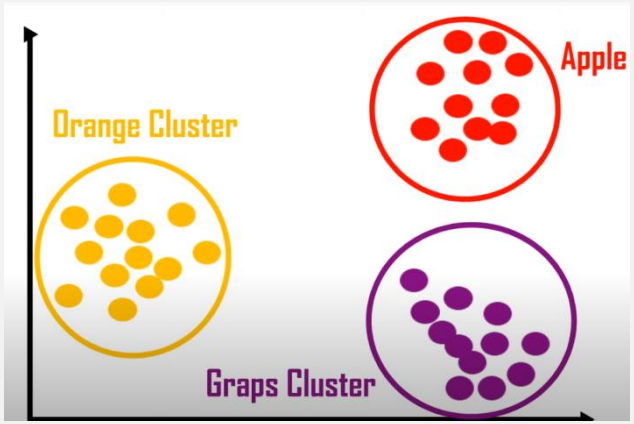
Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

Classification Vs. Regression

- Both classification and regression are supervised learning techniques.
- Key difference is type of output:
  - Continuous or Discrete
- Regression: Continuous
- Classification: Discrete

Classification Vs. Regression

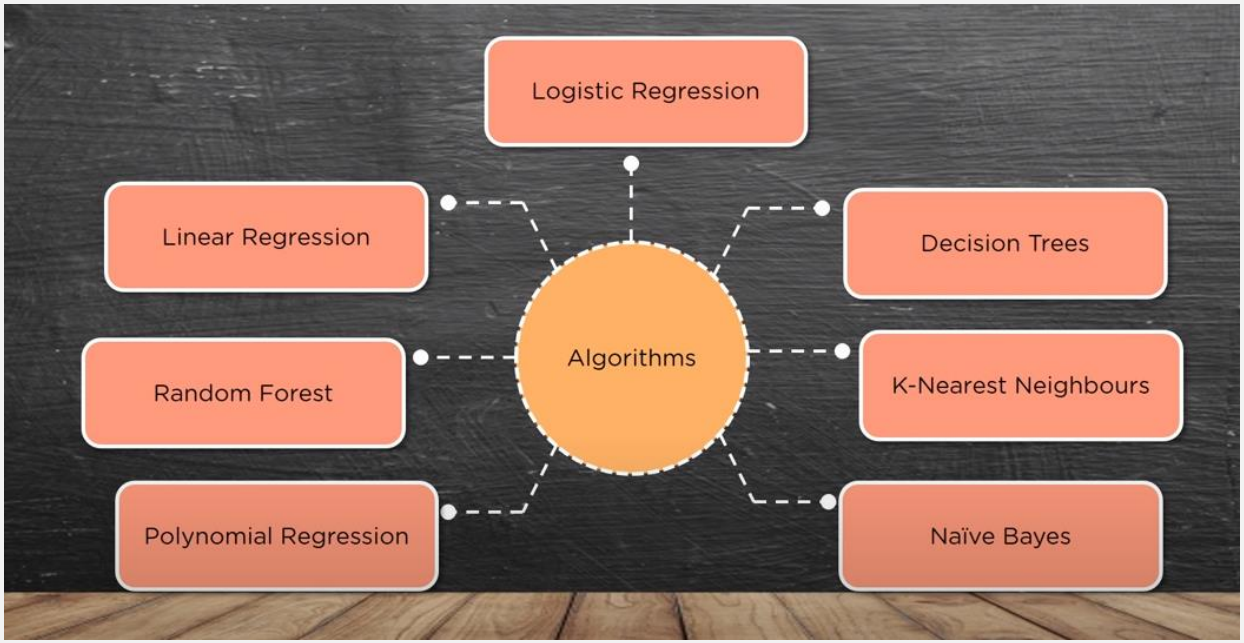




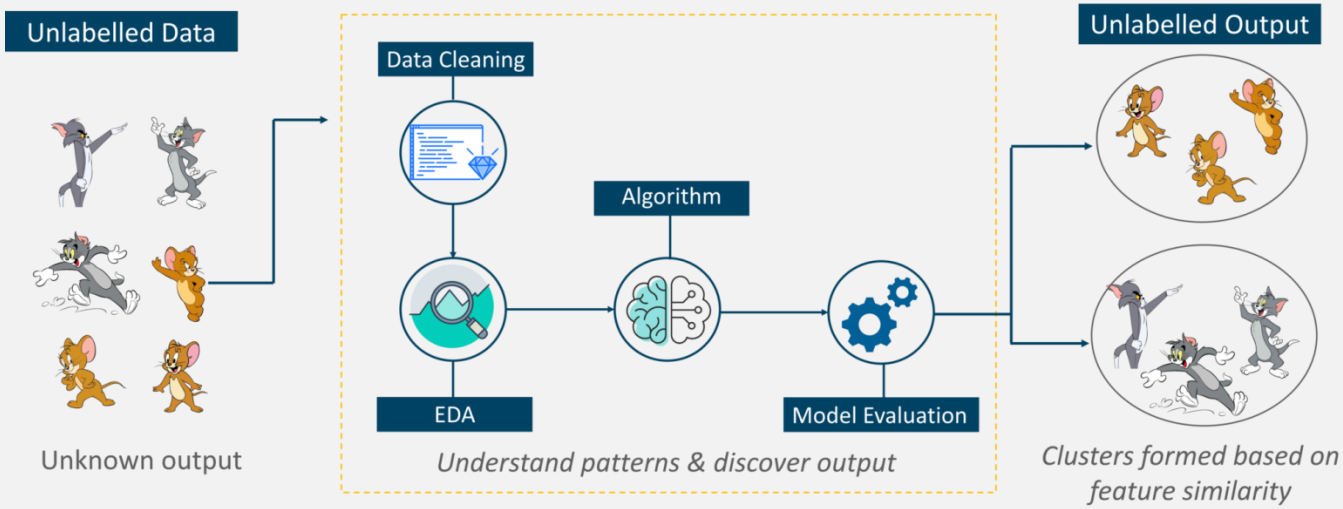
Supervised Learning

Here is the list of top algorithms currently being used for supervised learning are:

- **Polynomial regression**
- **Random forest**
- **Linear regression**
- **Logistic regression**
- **Decision trees**
- **K-nearest neighbors**
- **Naive Bayes**



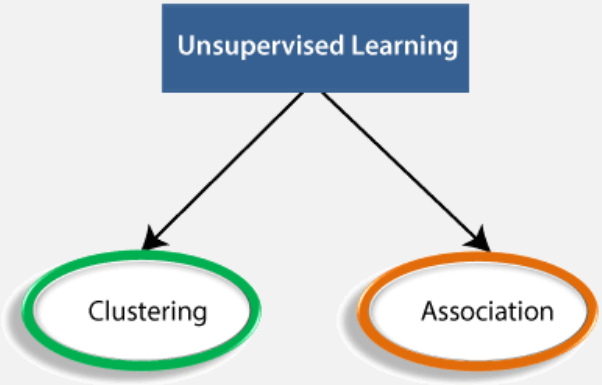
Unsupervised Learning



Unsupervised learning involves training by using unlabeled data and allowing the model to act on that information without guidance

In unsupervised learning, the training data is unknown and unlabeled - meaning that no one has looked at the data before

Unsupervised Learning

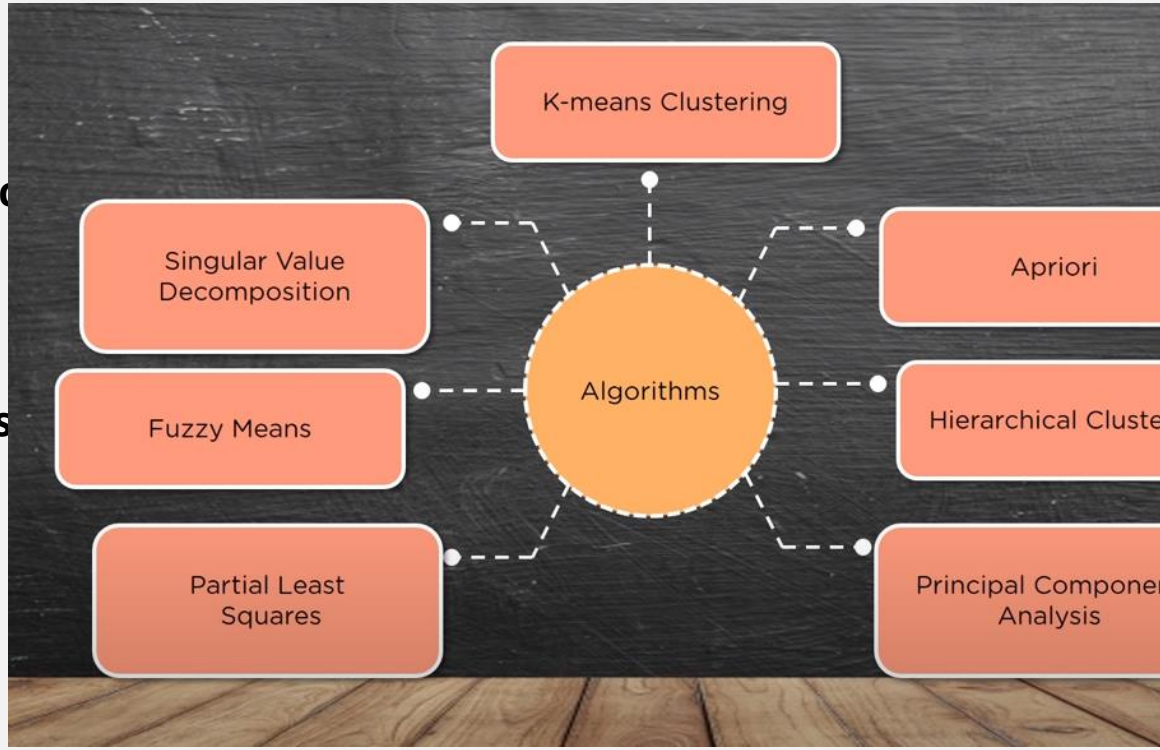


Clustering: Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.

Association: An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

Unsupervised Learning

- Partial least squares
- Fuzzy means
- Singular value decompositi
- K-means clustering
- Apriori
- Hierarchical clustering
- Principal component analys

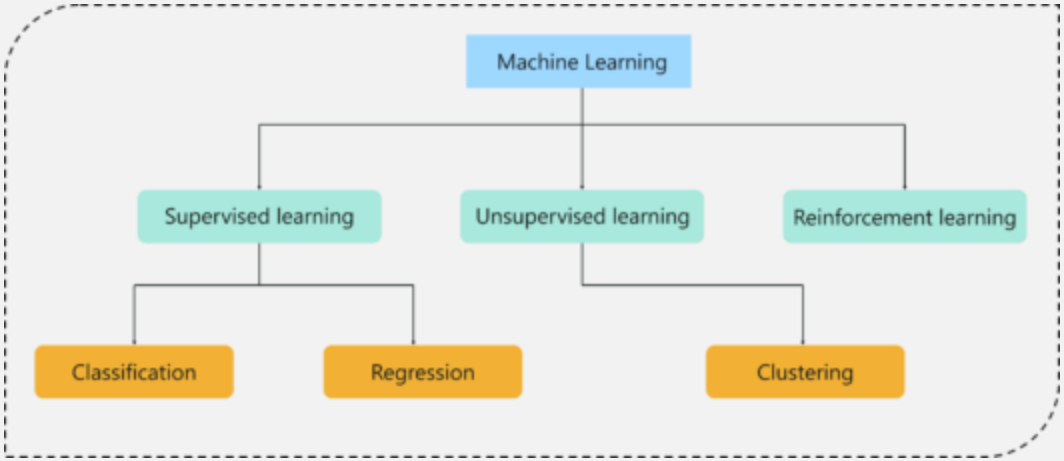


Reinforced Learning

Reinforcement Learning is a part of Machine learning where an agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the rewards which it gets from those action

Reinforcement Learning is mainly used in advanced Machine Learning areas such as self-driving cars, AplhaGo, etc.

Type Of Problems In Machine Learning

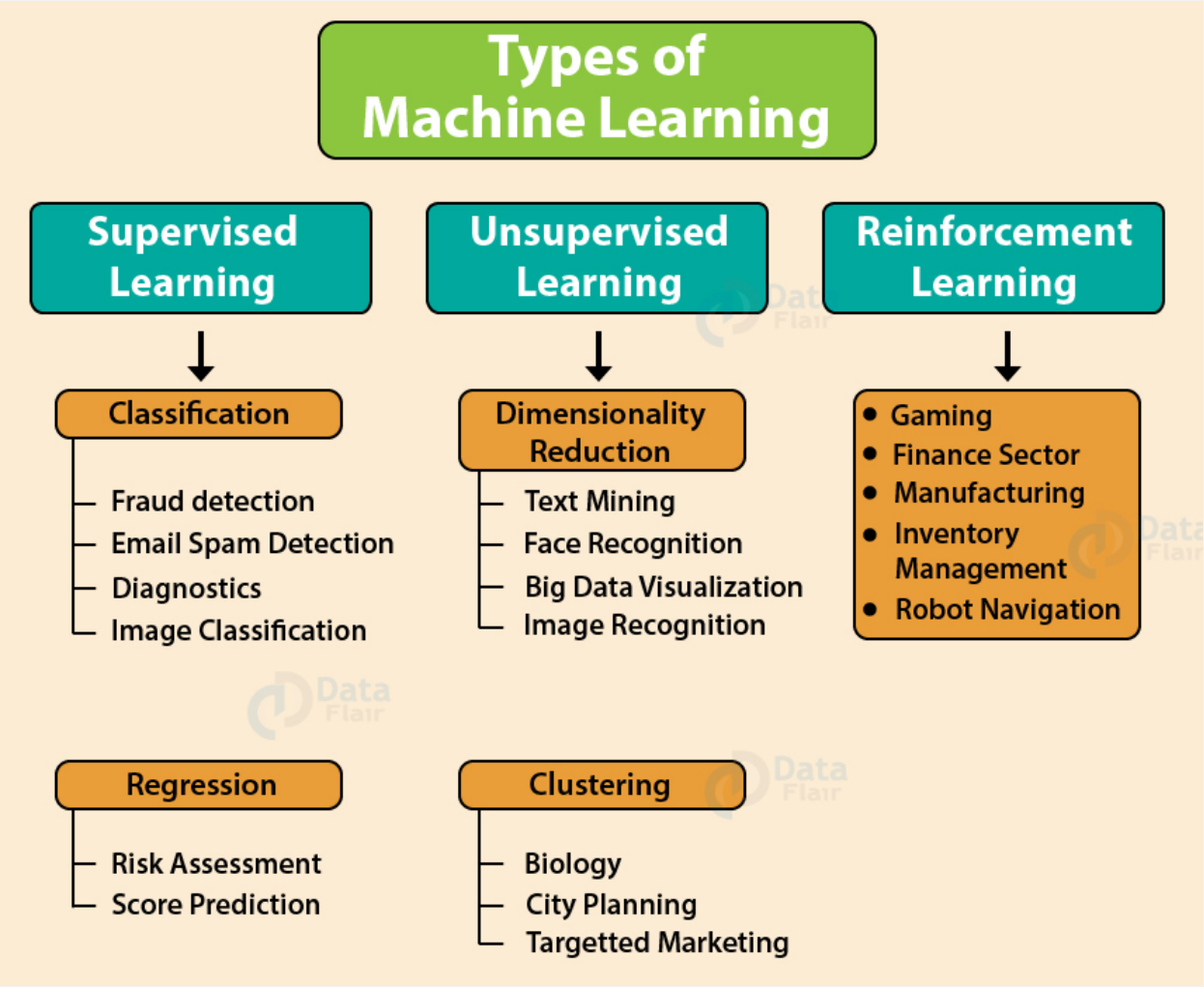


Regression: In this type of problem the output is a continuous quantity

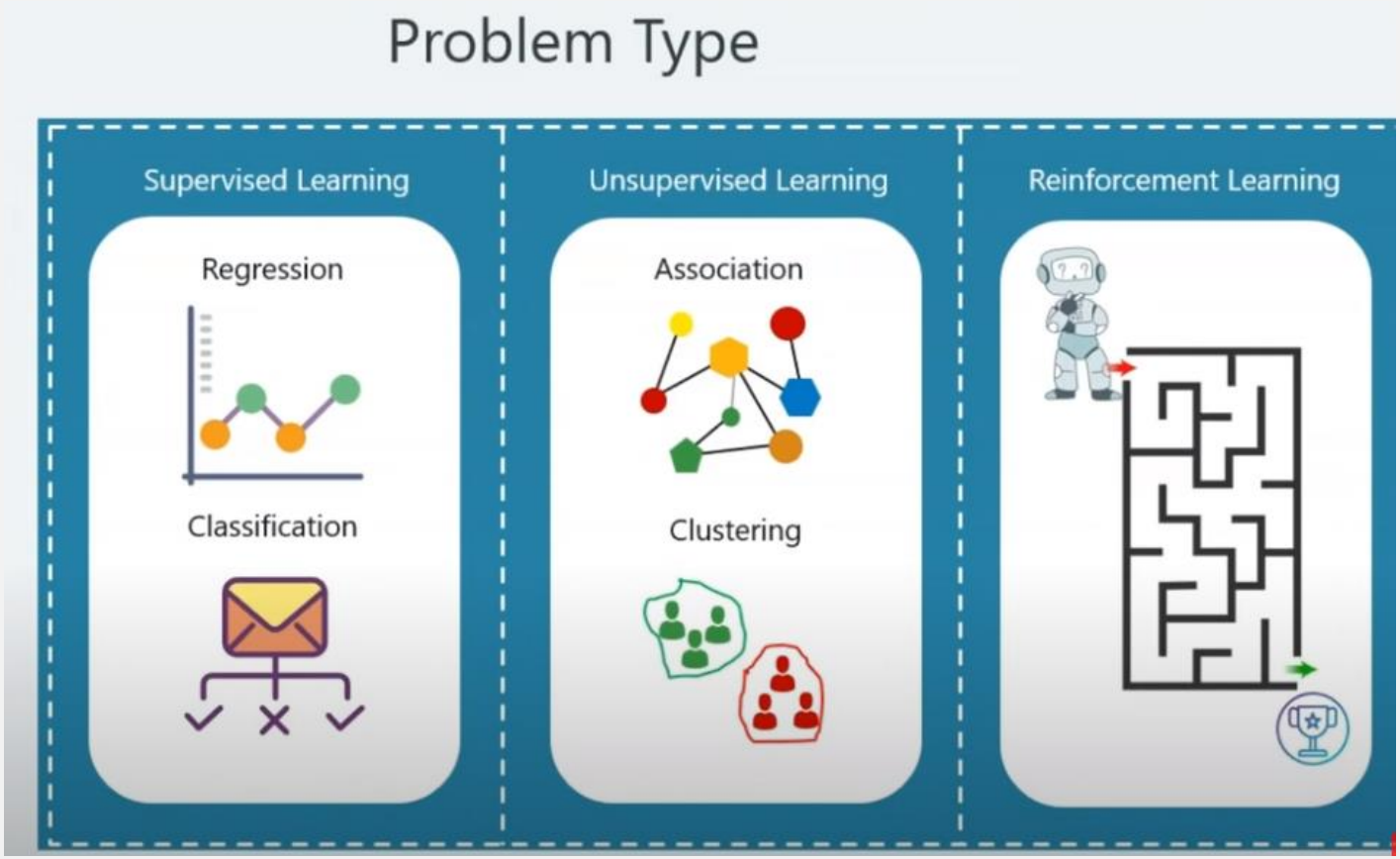
Classification: In this type, the output is a categorical value

Clustering: This type of problem involves assigning the input into two or more clusters based on feature similarity.

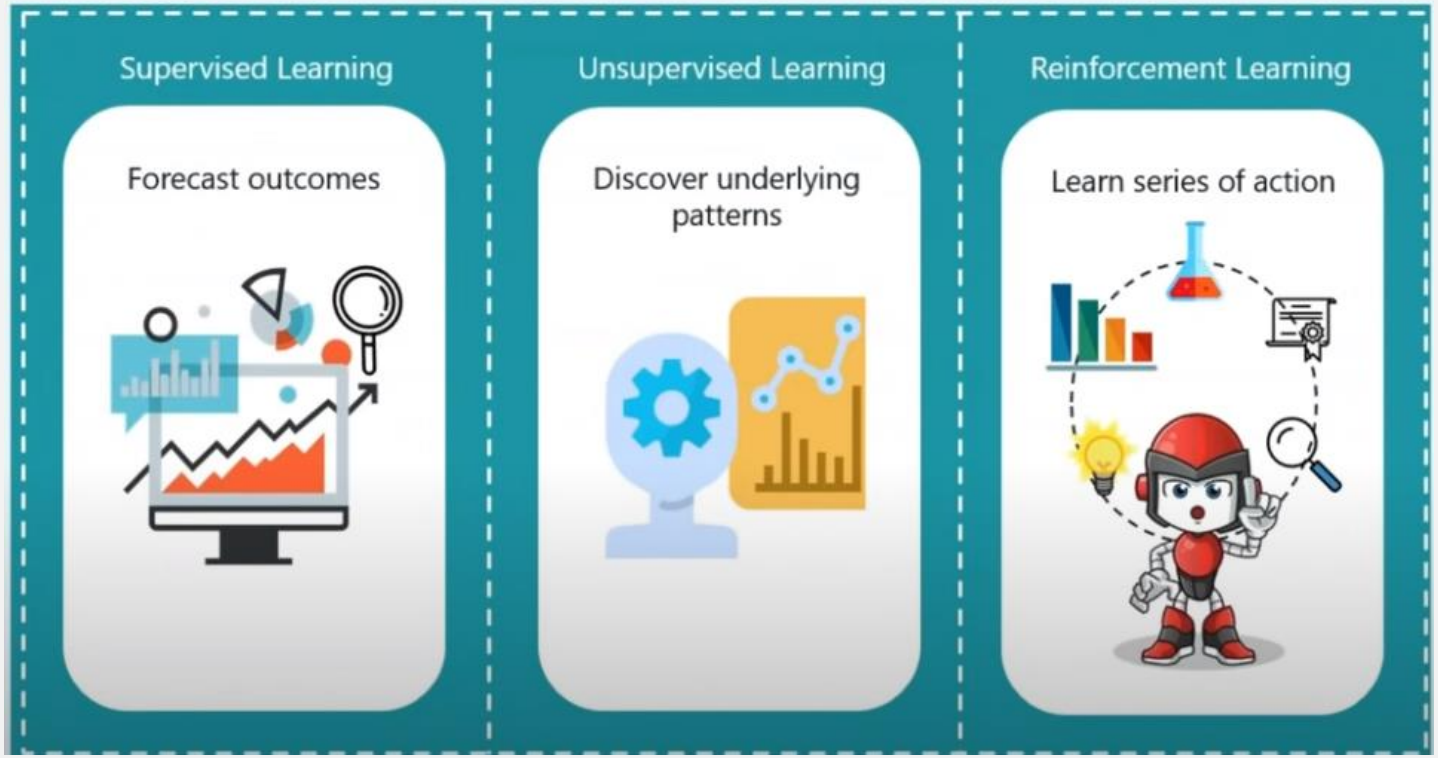
Type Of Problems In Machine Learning



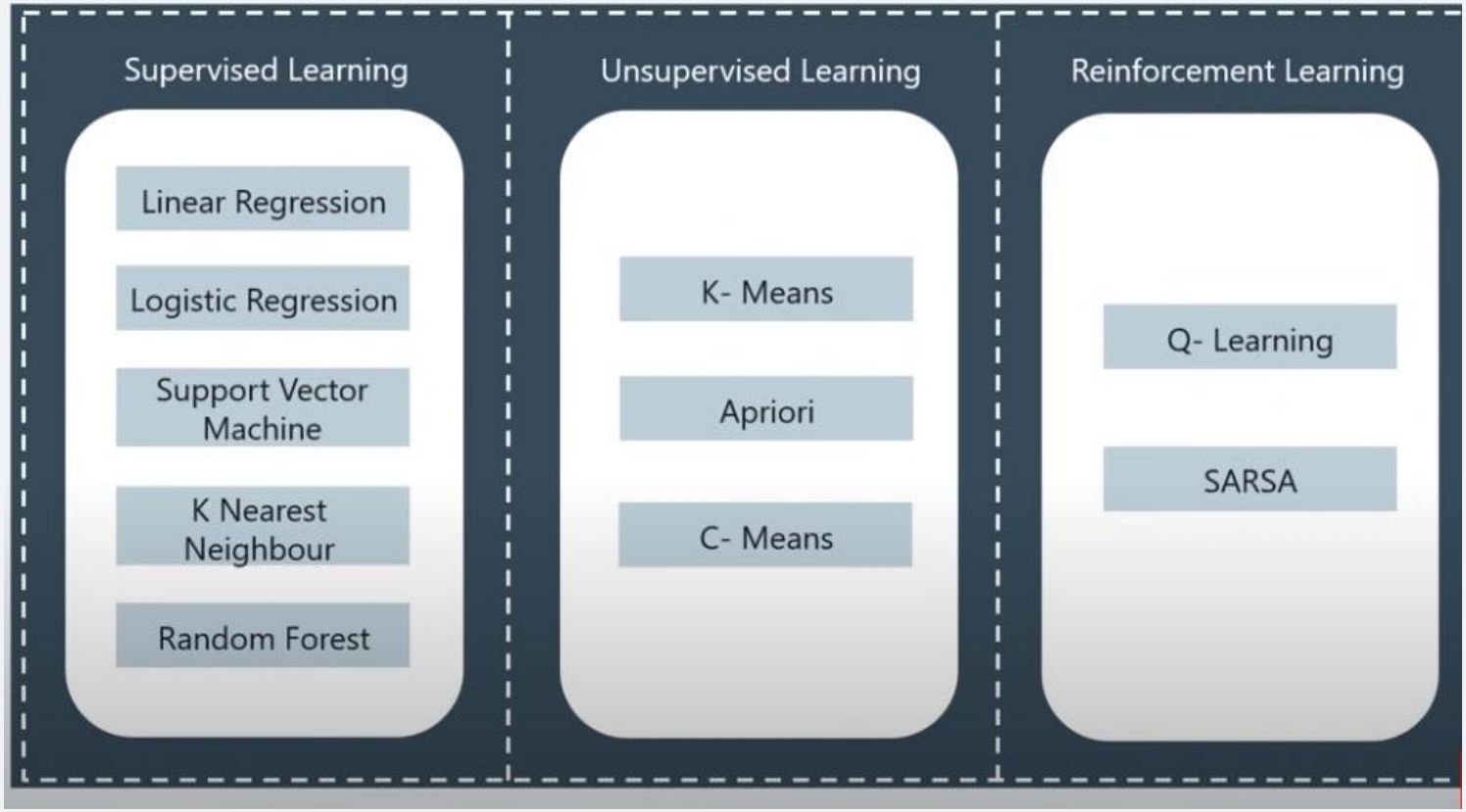
Problem Types



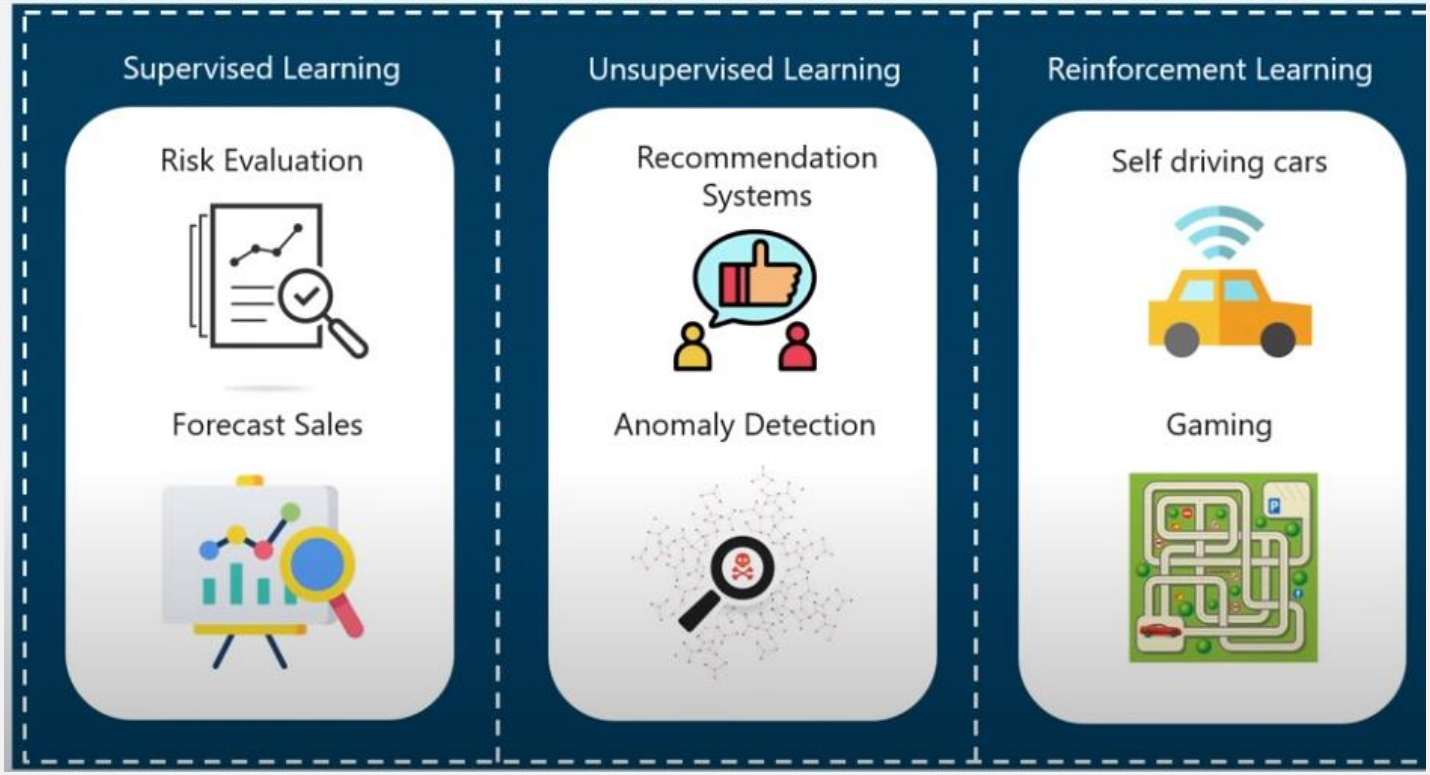
Aim


Supervised Learning
Forecast outcomes
Unsupervised Learning
Discover underlying patterns
Reinforcement Learning
Learn series of action

Popular Algorithms



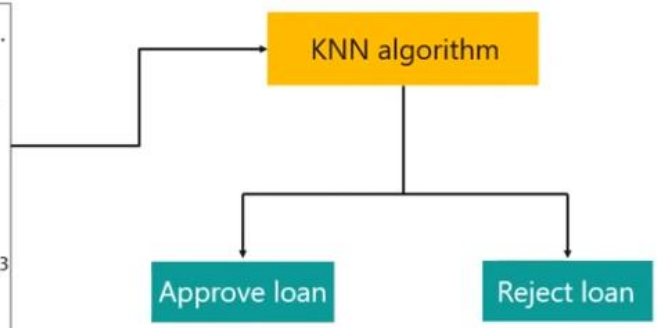
Applications



Use Case 1

Problem Statement: Study a bank credit dataset and make a decision about whether to approve the loan of an applicant based on his profile

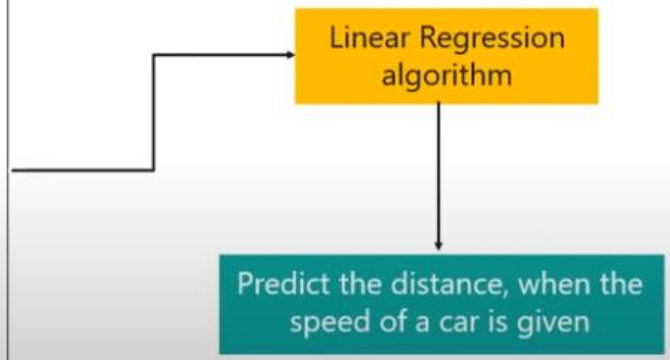
```
$ Account.Balance                 : int  1 1 2 1 1 1 1 1 4 2 ...
$ Duration.of.Credit..month.      : int  18 9 12 12 12 10 8 6 18 24 ..
$ Payment.Status.of.Previous.Credit: int  4 4 2 4 4 4 4 4 4 2 ...
$ Purpose                         : int  2 0 9 0 0 0 0 0 3 3 ...
$ Credit.Amount                   : int  1049 2799 841 2122 2171 2241
$ Value.Savings.Stocks            : int  1 1 2 1 1 1 1 1 1 3 ...
$ Length.of.current.employment    : int  2 3 4 3 3 2 4 2 1 1 ...
$ Instalment.per.cent             : int  4 2 2 3 4 1 1 2 4 1 ...
$ Sex...Marital.Status            : int  2 3 2 3 3 3 3 3 2 2 ...
$ Guarantors                      : int  1 1 1 1 1 1 1 1 1 1 ...
$ Duration.in.Current.address     : int  4 2 4 2 4 3 4 4 4 4 ...
$ Most.valuable.available.asset   : int  2 1 1 1 2 1 1 1 3 4 ...
$ Age..years.                     : int  21 36 23 39 38 48 39 40 65 23
$ Concurrent.Credits              : int  3 3 3 3 1 3 3 3 3 3 ...
$ Type.of.apartment               : int  1 1 1 1 2 1 2 2 2 1 ...
$ No.of.Credits.at.this.Bank      : int  1 2 1 2 2 2 2 1 2 1 ...
$ Occupation                      : int  3 3 2 2 2 2 2 2 1 1 ...
$ No.of.dependents                : int  1 2 1 2 1 2 1 2 1 1 ...
$ Telephone                       : int  1 1 1 1 1 1 1 1 1 1 ...
$ Foreign.Worker                  : int  1 1 1 2 2 2 2 2 1 1 ...
```



Use Case 2

Problem Statement: To establish a mathematical equation for **distance** as a function of **speed**, so you can use it to predict **distance** when only the **speed** of the car is known.

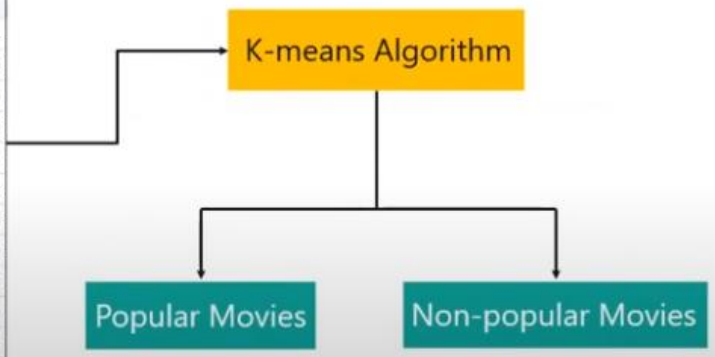
```
> cars
   speed dist
1      4    2
2      4   10
3      7    4
4      7   22
5      8   16
6      9   10
7     10   18
8     10   26
9     10   34
10    11   17
11    11   28
12    12   14
13    12   20
14    12   24
15    12   28
```



Use Case 3

Problem Statement: To cluster a set of movies as either good or average based on their social media out reach

	director_facebook_likes	actor_3_facebook_likes	actor_1_facebook_likes	cast_total_facebook_likes
AvatarÂ	0	855	1000	4834
Pirates of the C...	563	1000	40000	48350
SpectreÂ	0	161	11000	11700
The Dark Knigh...	22000	23000	27000	106759
John CarterÂ	475	530	640	1873
Spider-Man 3Â	0	4000	24000	46055
TangledÂ	15	284	799	2036
Avengers: Age ...	0	19000	26000	92000
Harry Potter an...	282	10000	25000	58753
Batman v Super...	0	2000	15000	24450
Superman Retur...	0	903	18000	29991
Quantum of Sol...	395	393	451	2023
Pirates of the C...	563	1000	40000	48486



Use Case 4

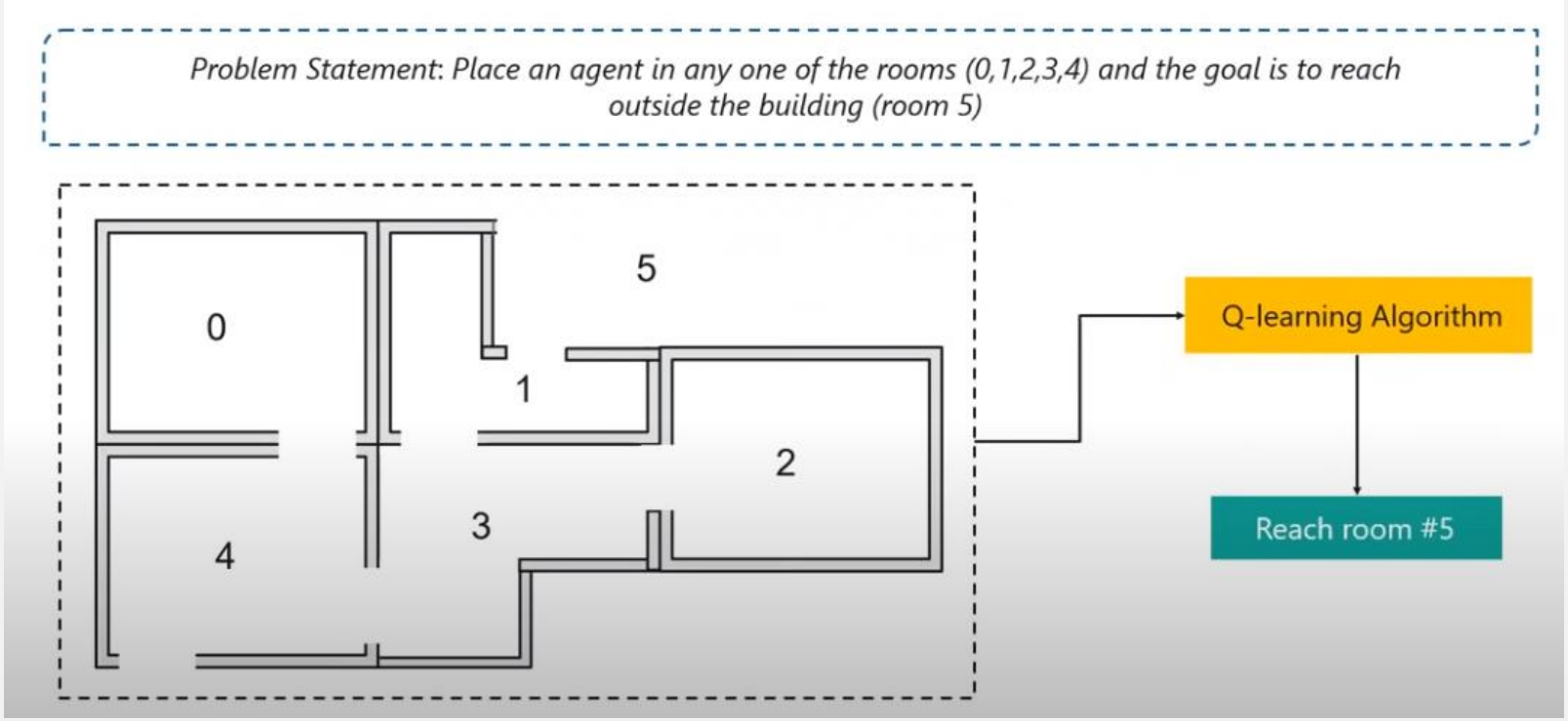
Problem Statement: To perform Market Basket Analysis by finding association between items bought at the grocery store

	A	B	C	D	E	F	G	H
1	citrus fruit	semi-finisl	margarine	ready soups				
2	tropical fr	yogurt	coffee					
3	whole milk							
4	pip fruit	yogurt	cream che	meat spreads				
5	other vege	whole mill	condensec	long life bakery product				
6	whole mill	butter	yogurt	rice	abrasive cleaner			
7	rolls/buns							
8	other vege	UHT-milk	rolls/buns	bottled be	liquor (appetizer)			
9	pot plants							
10	whole mill	cereals						
11	tropical fr	other vege	white brea	bottled wa	chocolate			
12	citrus fruit	tropical fr	whole mill	butter	curd	yogurt	flour	bottled wa
13	beef							
14	frankfurte	rolls/buns	soda					
15	chicken	tropical fruit						

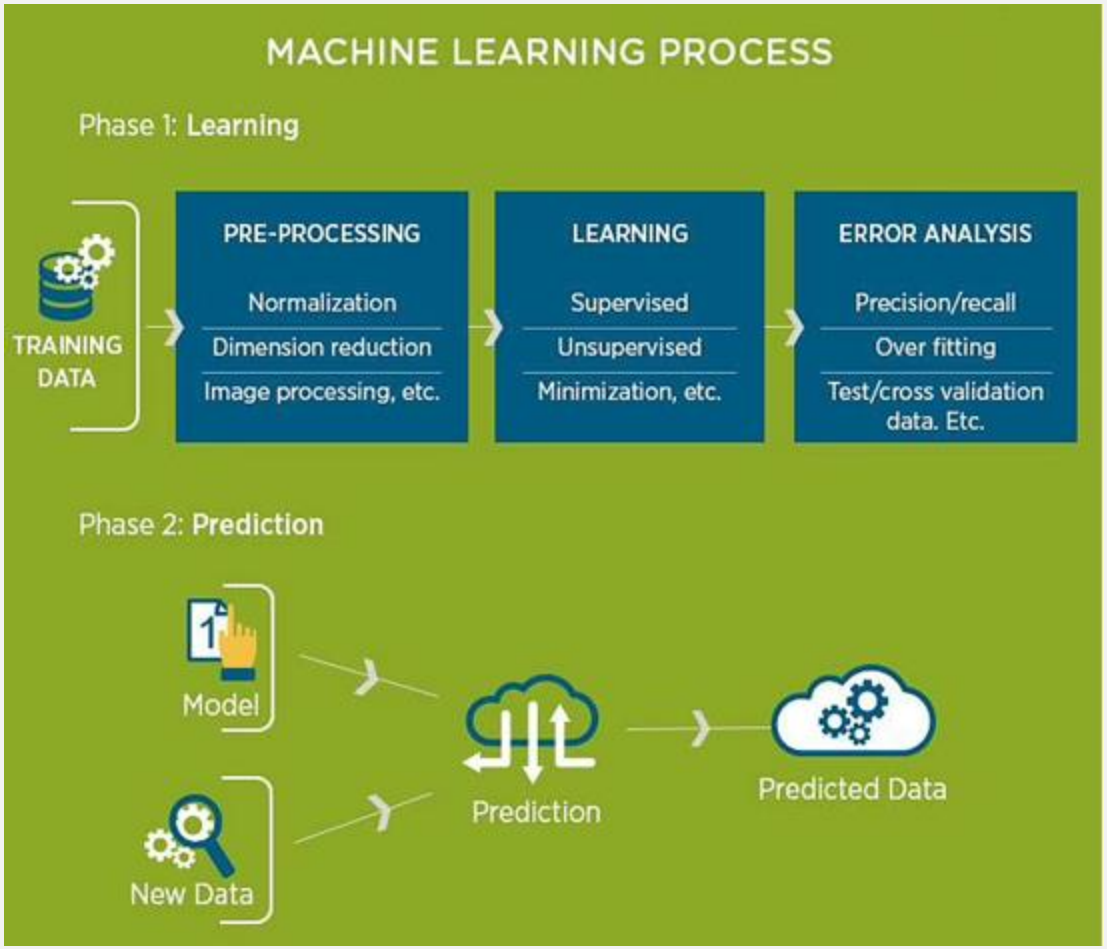
Association rule mining & Apriori Algorithm

Perform Market Basket Analysis

Use Case 5


Problem Statement: Place an agent in any one of the rooms (0,1,2,3,4) and the goal is to reach outside the building (room 5)
0
1
2
3
4
5
Q-learning Algorithm
Reach room #5

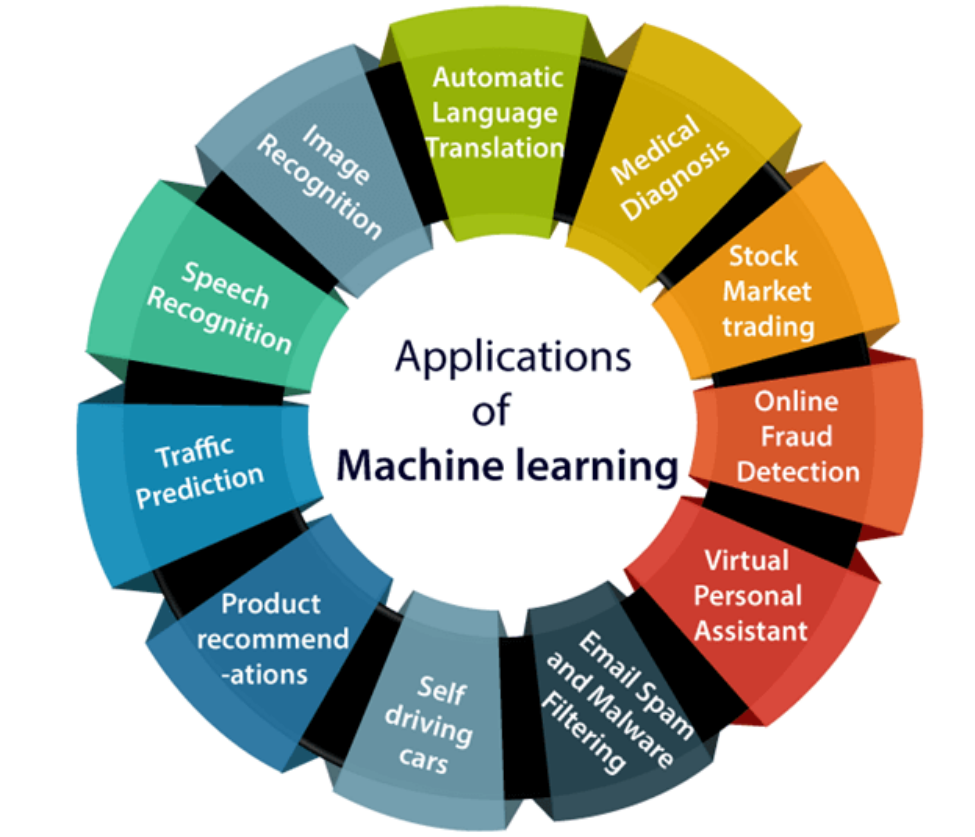
Machine Learning Process



Top 10 Machine Learning Algorithms

- Linear regression
- Logistic regression
- Decision tree
- SVM algorithm
- Naive Bayes algorithm
- KNN algorithm
- K-means
- Random forest algorithm
- Dimensionality reduction algorithms
- Gradient boosting algorithm and AdaBoosting algorithm

Machine Learning Applications



Machine Learning Applications

Recommendation Systems

Used in almost all domains involving

- **Netflix** – AI based system
- **Amazon** – Product purchase recomm
- **Spotify** – Music and playlist generati
- **Facebook** – Customized content on feed

Machine Learning Applications

Self Driving Cars

Powered by Machine Learning

- Machine learning and IoT work hand in hand.
- Powerful algorithms used
- Reaction time is critical

Machine Learning Applications

Social Media

Highly customized and efficient user experience

- Face recognition
- Customized Advertisements
- Feed enhancements

Machine Learning Applications

Sentiment Analysis

Vital if you sell products/services

- Used to determine the opinion of users.
- Assesses emotion and tone of the message.
- Gives way for better product development.

Machine Learning Applications

Healthcare

Machine Learning is revolutionizing healthcare

- Helping in identifying and diagnosing diseases.
- Key part in drug manufacturing research.
- Analysis of medical imaging products
- Pandemic outbreak prediction

Machine Learning Applications

Banking Sector

An important application of Machine Learning

- Fraud detection and prevention
- Portfolio management tools
- Network security protocols
- Personalized assistants

Machine Learning Applications

Traffic Prediction

Helping immensely to reduce costs and pollution

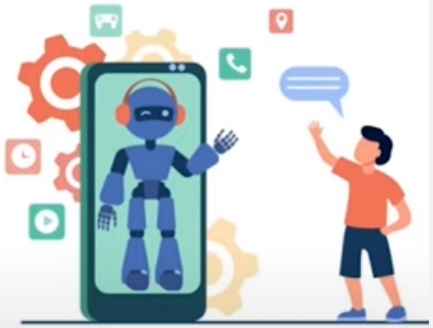
- Real-time traffic assessment and recommendation
- Automated routing based on traffic conditions
- Traffic information fed directly to the car's dashboard

Machine Learning Applications

Virtual Assistants

Adding a lot of convenience to life

- **Apple** – Siri
- **Amazon** – Alexa
- **Google** – Google Assistant
- **Microsoft** - Cortana



Machine Learning Applications

Language Translation

Imagine if you went on a vacation..

- Concepts of Natural Language Processing (NLP) are used.
- Integration of Augmented Reality (AR) is now popular.
- Accuracy rate is exponentially better than before.

THANK YOU !!!

Amol Patil - 9822291613