Types of Machine Learning Models and Framework

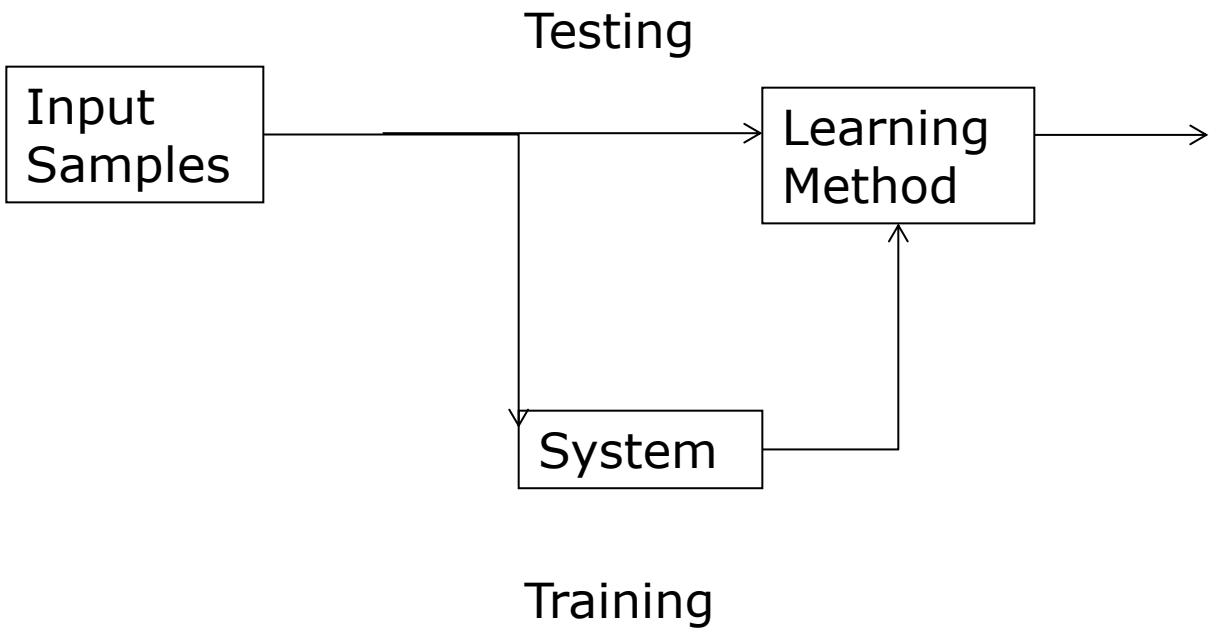
By

Prof(Dr) Premanand P Ghadekar

Outline

- Types of Machine Learning Models
- Machine Learning Framework
- Machine Learning Model Development and Applications

Learning System Model



Process Involved in Machine Learning

- Data Gathering-Sufficient Historical data
- Data Preprocessing-Raw data need to preprocess
- Choose the Model
- Train Model
- Test model
- Tune Model
- Prediction

Introduction

- The human brain has a unique capability called categorization. **Categorization** is the ability to assign label(or pattern class) to as object.
- The process of assigning a meaningful label to an unknown object is called as **Recognition**.
- **Example-**When we encounter an animal in a picture or in reality, we register its identity using its features. When an unknown animal is encountered, we try to recognize it by comparing its features (called) pattern) with known stored pattern that we already have.
- This process of comparing an unknown object with stored patterns to recognize the unknown object is called **Classification.**
- Thus, **Classification** is the process of applying a label or pattern class to an unknown instance.

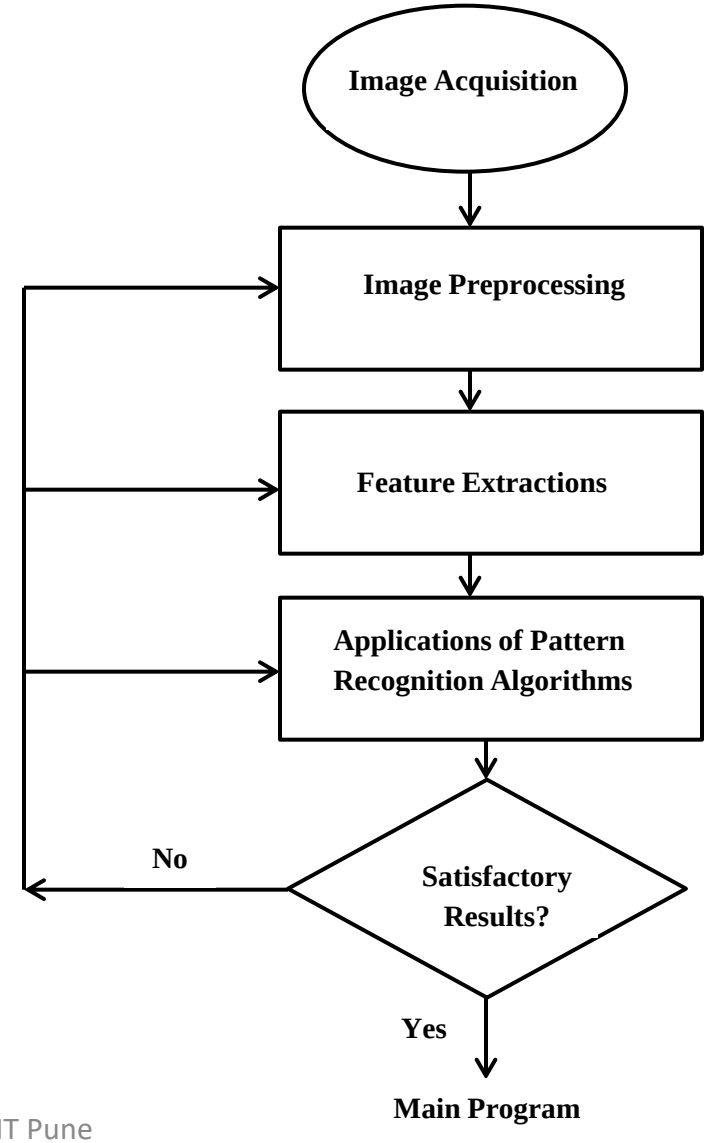
Object Recognition Process Steps

- In the absence of any prior knowledge of the object or stored pattern, we use a trial and error process to recognize the object.
- This trial and error process of grouping of objects is called **Clustering.**

- **A Feature vector is typically of the form**

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$$

PCA Algorithm



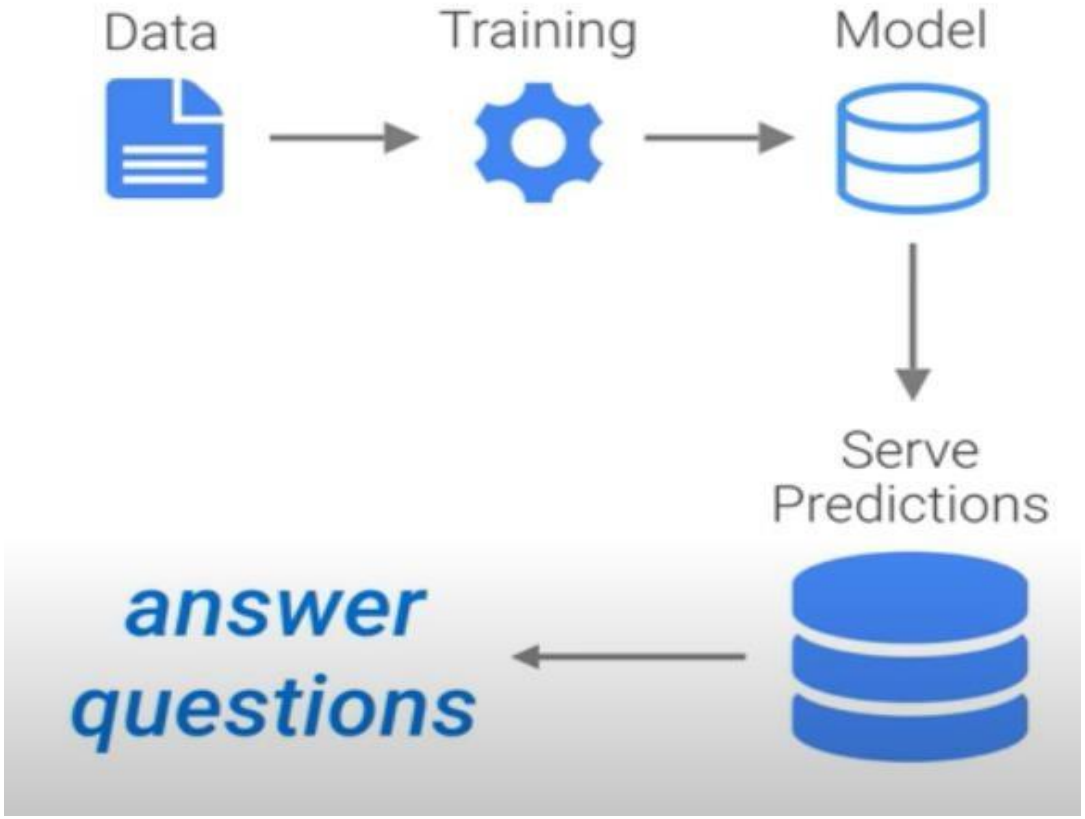
Algorithms

- The success of **Machine Learning** system also depends on the **Algorithms**.
- The algorithms control the search to find and build the knowledge structures.
- The learning algorithms should extract useful information from training examples.

What's Machine Learning?

- In Simplest way it can be defined as:

"using data to answer questions"



What's Machine Learning?

Categorized as:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Different types of Machine Learning Algorithms

- **Supervised Learning -Labelled Data**
- Learn an input and output map (Fraud Detection, Weather Forecasting etc.)
  - Prediction
  - Classification (discrete labels)-Categorical Output,
  - Regression (real values)-Continuous output
- **Unsupervised Learning -Unlabeled Data**
- Discover patterns in the data (Face recognition, text mining, City Planning etc.)
  - Clustering –Cohesive Grouping
  - Association-Frequent Co-occurrence
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction
- **Semi-supervised Learning-**It combines a small amount of labeled data with a large amount of unlabeled data during training (Text document classifier)
- **Reinforcement learning**
  - Decision making (Robot, Chess machine, Gaming, Stock Trading etc.)

Supervised Learning

- **Supervised Learning** occurs when an algorithm learns from example data and associated target responses that can consist of numeric values or string labels, such as classes or tags, in order to later predict the correct response when posed with new examples.
- The supervised approach is indeed similar to human learning under the supervision of a teacher.
- The teacher provides good examples for the student to memorize, and the student then derives general rules from these specific examples.
- The objective of a supervised classifier is to assign a class to an example after having examined some characteristics of the example itself (Features- Quantitative (numeric values) or Qualitative (string labels)).

Supervised Learning

A simple model of Supervised Learning is provided as given below-

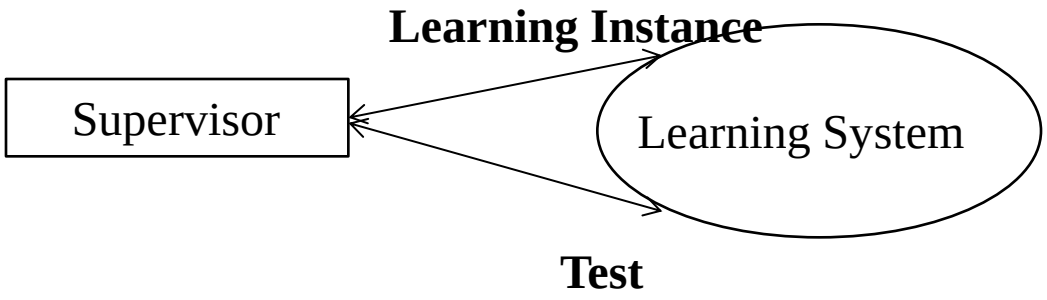
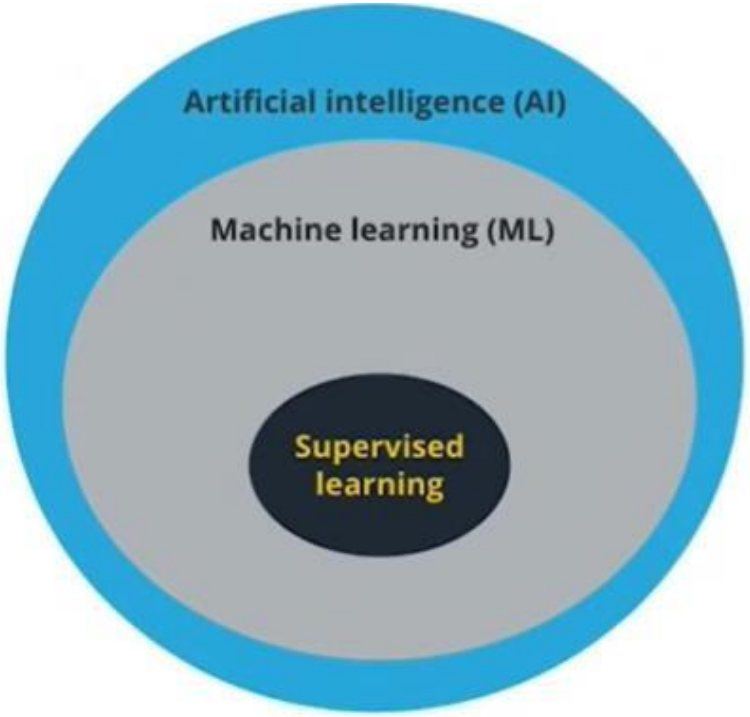
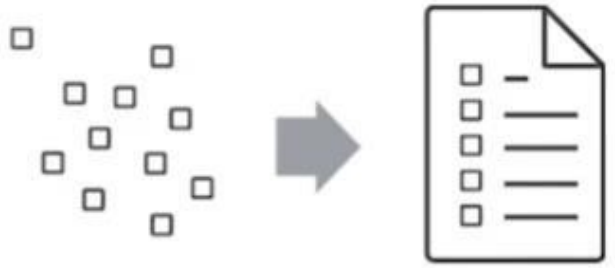


Fig-Supervisory System

- The Supervisor provides a label/cost for each pattern in a training set.
- Based on the training set, the system learns and generates a concept to classify the pattern.
- Once the system becomes a learnt system, the supervisor supplies the test data, using which it tests the system.
- This kind of learning, where there is an interaction between the Supervisor and the Learning system, is called **Supervised Learning.**

Supervised Learning

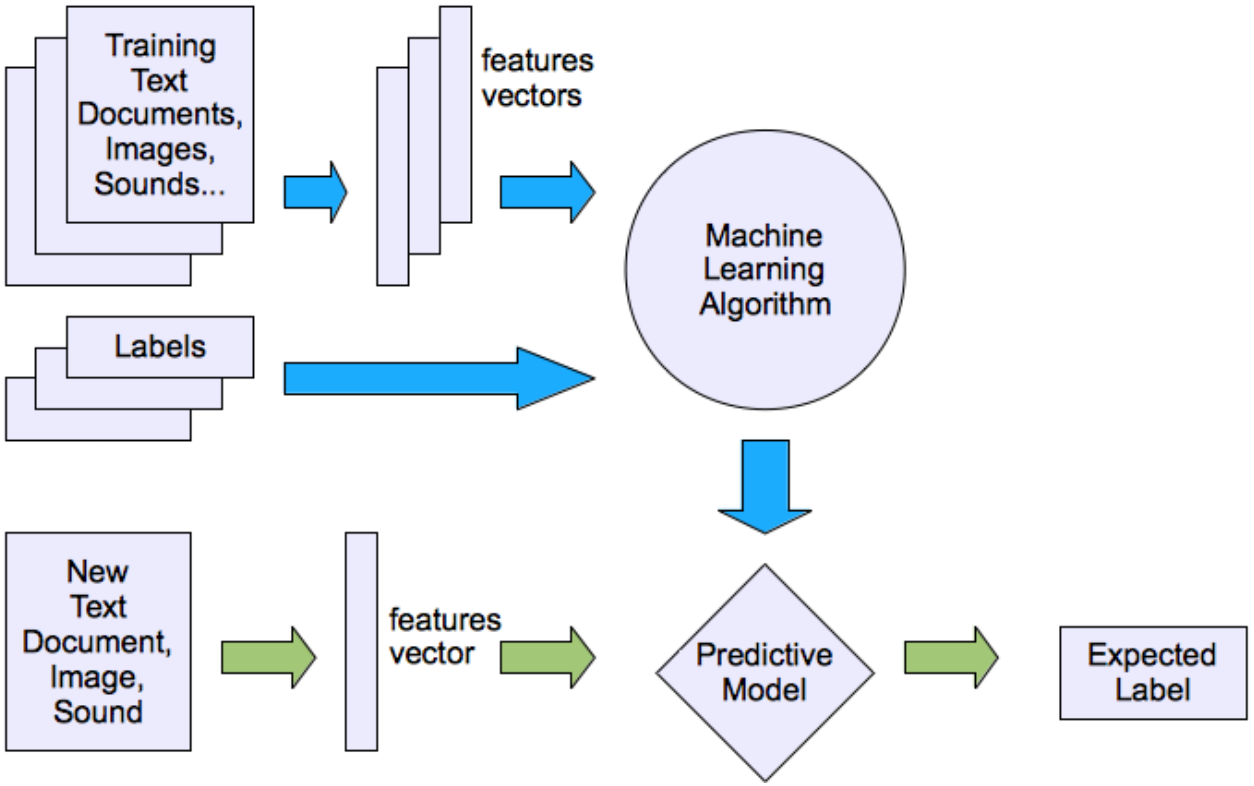




Example-driven training; every data point has a corresponding **label**

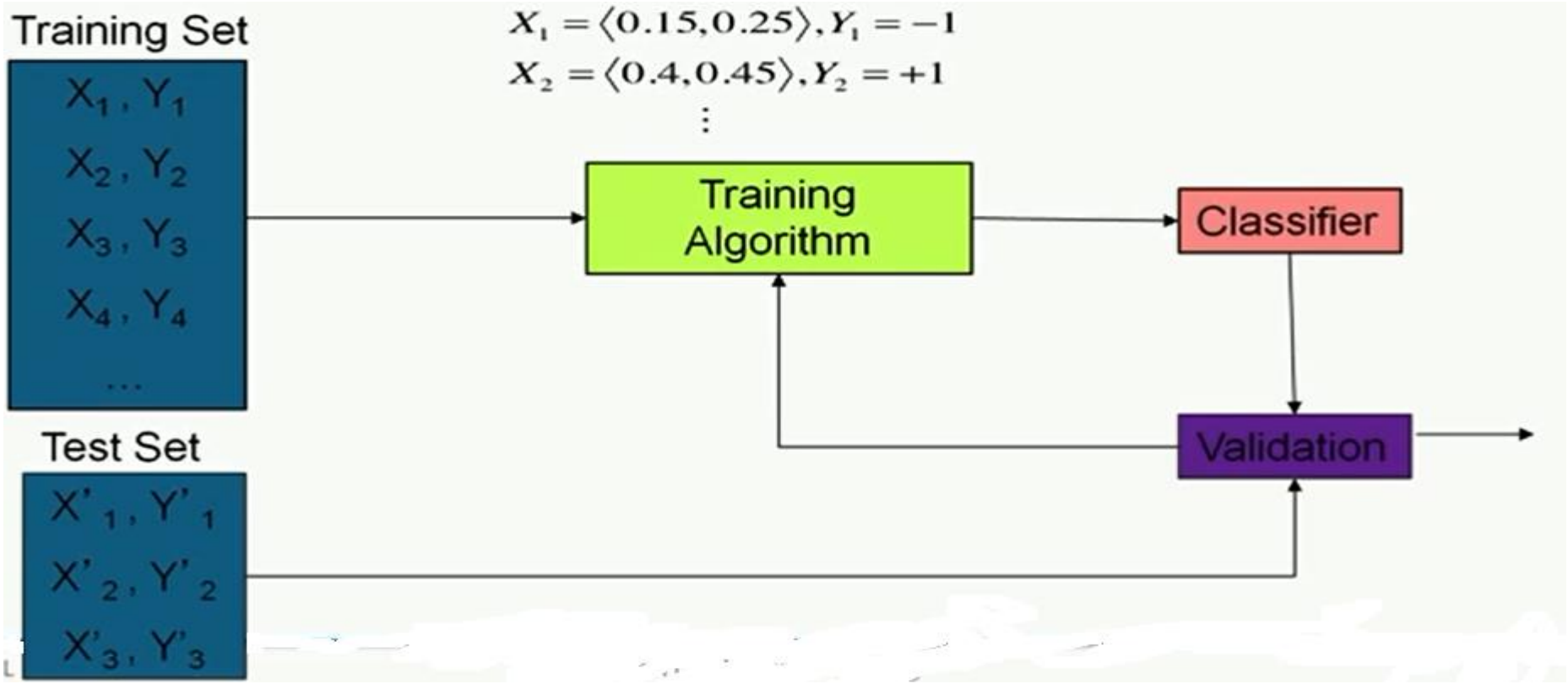
Supervised Learning

- **Supervised Learning**



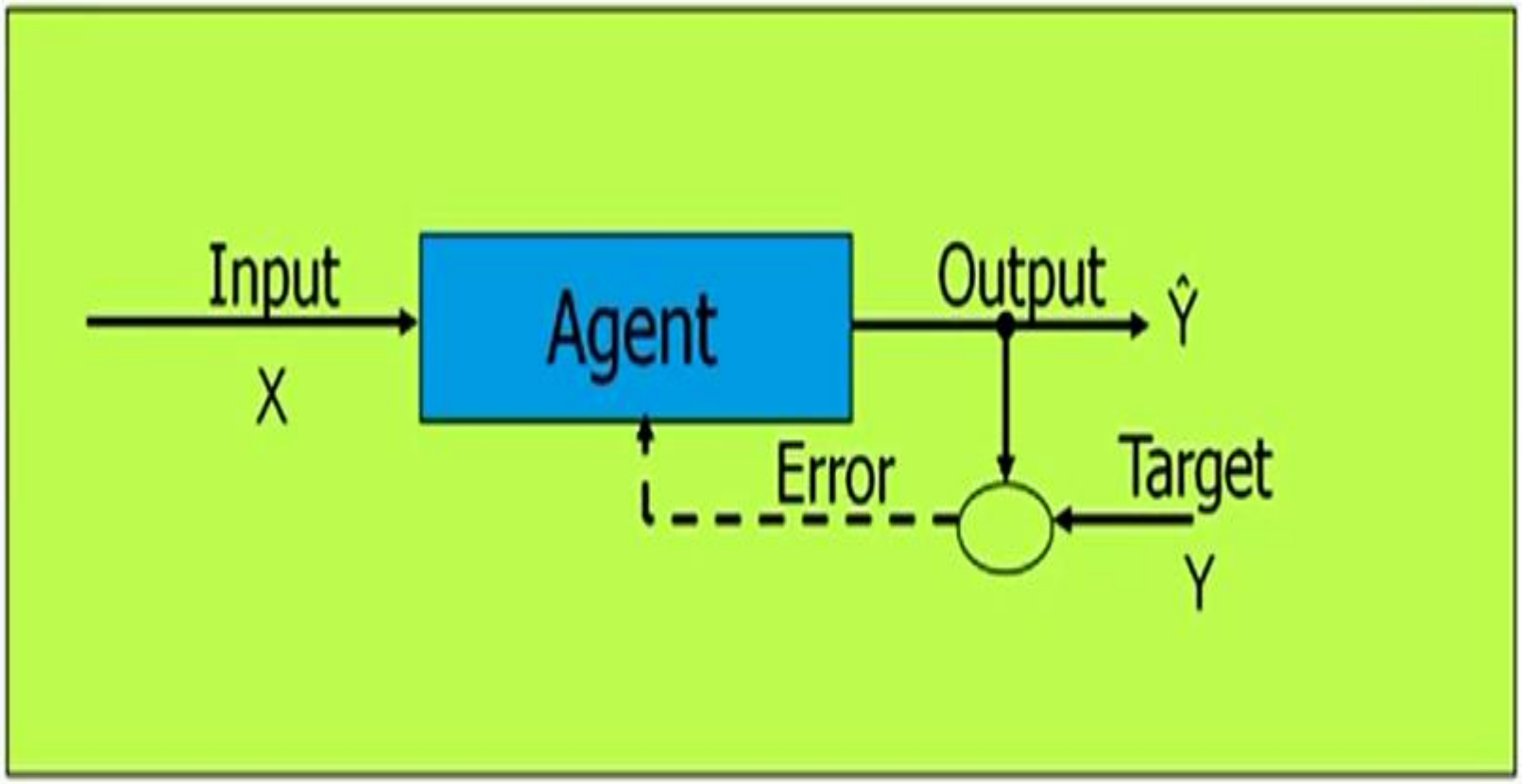
The Process

Training Dataset: The sample of data used to fit the model.
Test Dataset: The sample of data used to provide an unbiased evaluation of a final model fit on the **Training Dataset.**



Training

Training a Model simply means **Learning** (determining) good values for all the weights and the bias from labeled examples.



In supervised **learning**, a **machine learning** algorithm builds a model by examining many examples and attempting to find a model that minimizes loss; this process is called **empirical risk minimization.**

Applications

- **Credit Card Fraud Detection**
  - Valid transaction or not
- **Sentiment Analysis**
  - Opinion Mining, Buzz Analysis etc.
- **Churn Prediction**
  - Potential Churner or not
- **Medical Diagnosis**
  - Risk Analysis
- **Face Recognition**
- **Image Classifications**

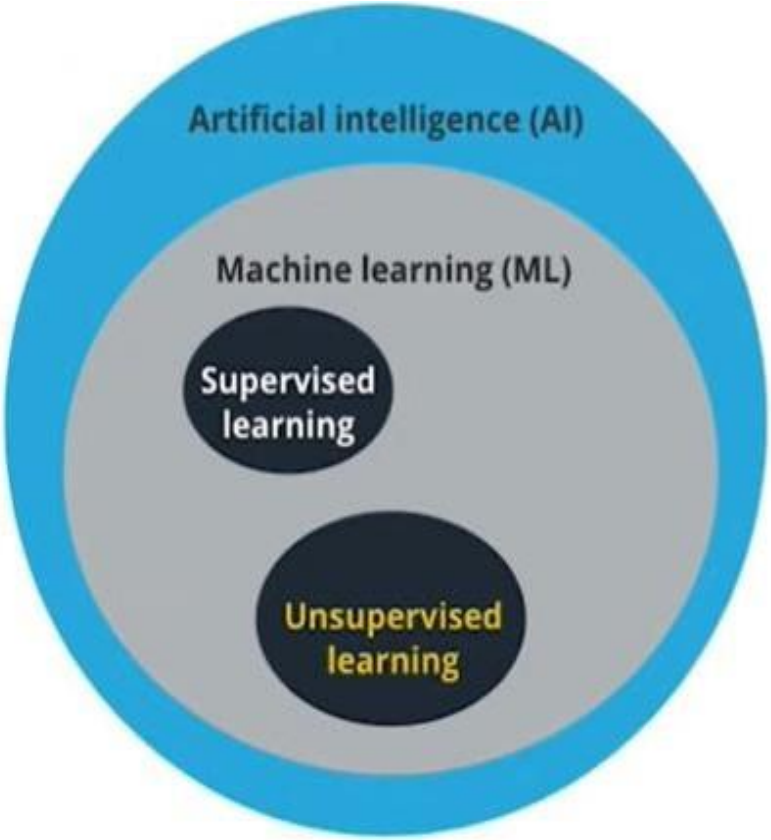
Unsupervised Learning

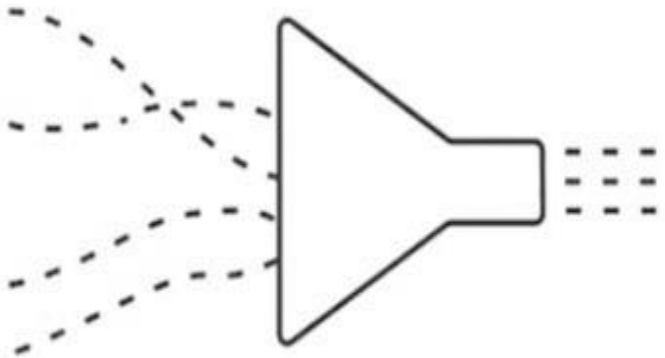
- **Unsupervised Learning** is a type of **Machine Learning** algorithm used to draw inferences from datasets consisting of input data without labeled responses.
- **Unsupervised Learning** occurs when an algorithm learns from plain examples without any associated response, leaving to the algorithm to determine the data patterns on its own.
- It resembles the methods humans use to figure out that certain objects or events are from the same class, such as by observing the degree of similarity between objects.

Unsupervised Learning

- **Unsupervised Learning**
  - There is no explicit Supervisor component in an Unsupervised system.
  - The learning system itself learns by trial and error method.
  - The instances themselves, based on similarity measure, form groups or clusters.
  - The goal of **Clustering** is similar to that of **Classification.**
  - However, it is performed when domain knowledge is not available.
  - **Example**
  - When we encounter an animal that we have not seen earlier, we often use trial and error to group the unknown animal with the closest animal familiar to us.
  - This process is called **Natural Grouping**

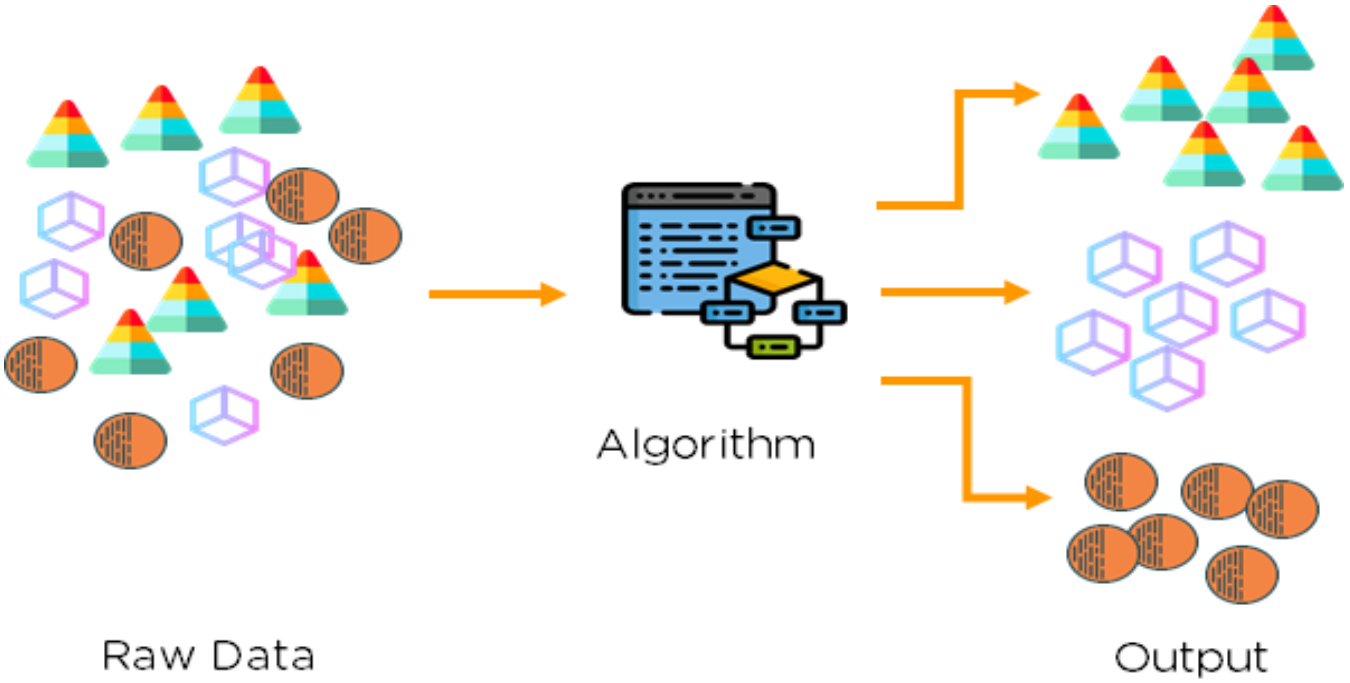
Unsupervised Learning





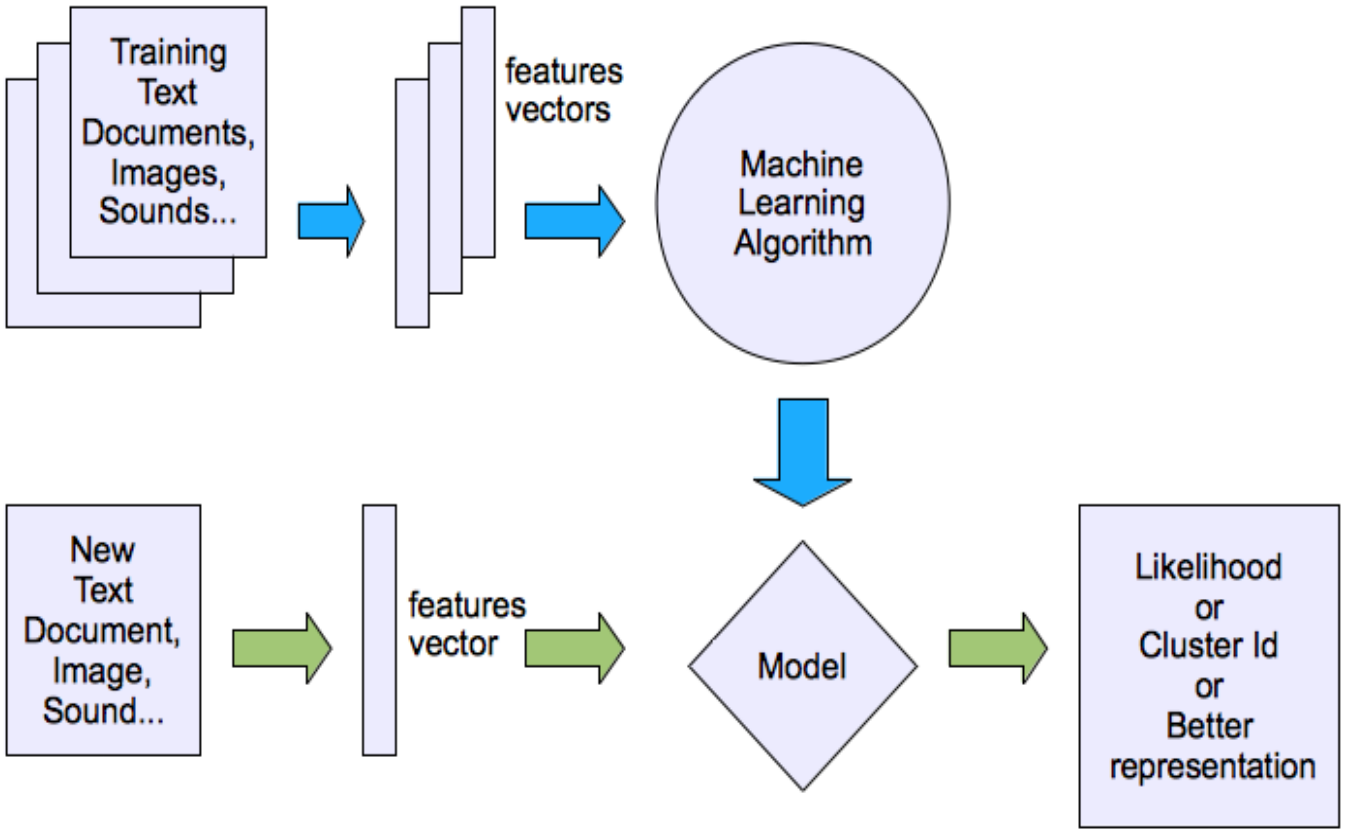
Patterns are discovered in unlabeled data

Unsupervised Learning

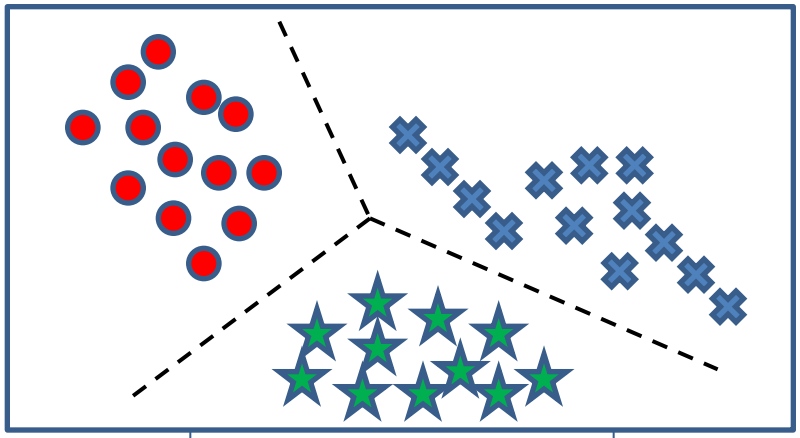


Machine learning structure

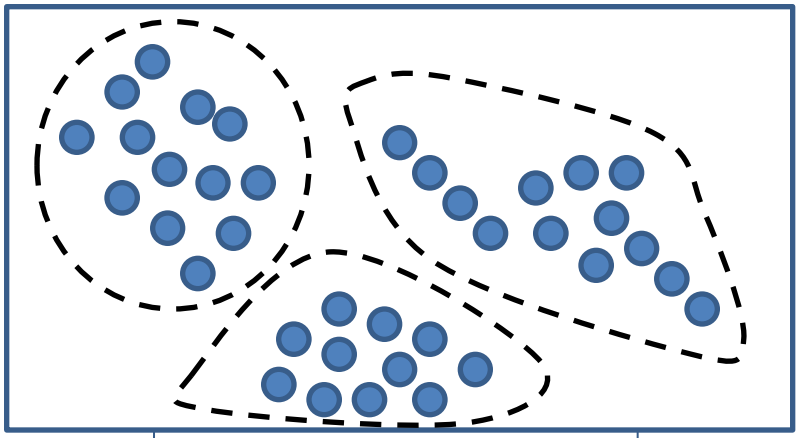
❖ Unsupervised Learning



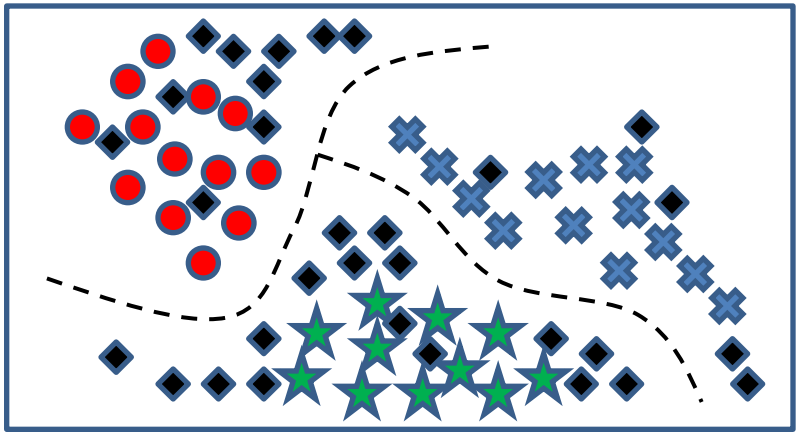
Algorithms



Supervised Learning



Unsupervised Learning



Semi-Supervised Learning

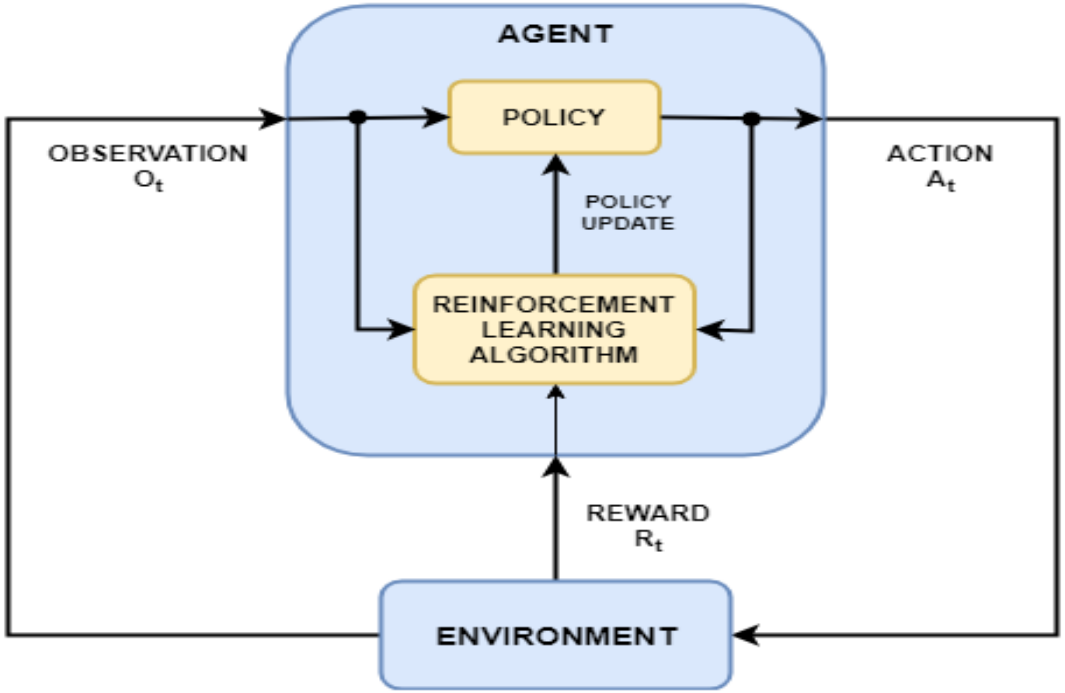
Supervised v/s Unsupervised

- In **Supervised learning**, you train the machine using data which is well **"labeled."**
- **Unsupervised learning** is a machine learning technique, where you **do not need** to **supervise the model.**
- **Supervised learning** allows you to collect data or produce a data output from **the previous experience.**
- **Unsupervised machine learning helps** you to finds all kind of **unknown patterns in data.**
- **Regression and Classification** are **two types of supervised machine learning techniques.**
- **Clustering and Association** are two types of **Unsupervised learning**.
- In a **supervised learning model, input and output variables** will be given while with **unsupervised learning model, only input data will be given**

Reinforced Learning

❖ Reinforced Learning

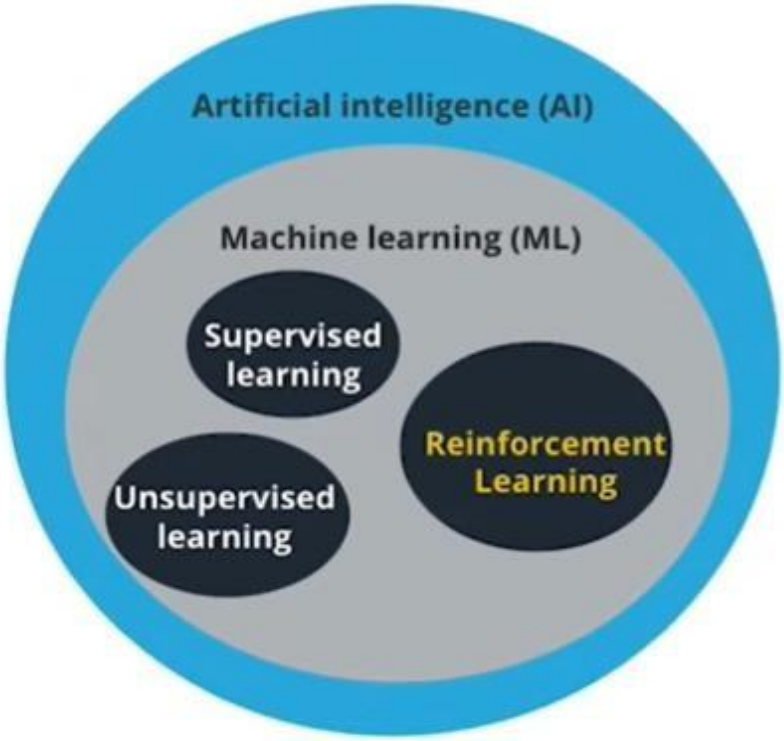
- In this method, the output of the learning system is binary.
- The binary feedback of right or wrong is sent back to the input and is used to reinforce learning from the data.
- The role of reinforced learning is that of a critic who is authorized to say right or wrong.
- Thus the learning continues till the critic agrees with the learning system.



Reinforced Learning

- **Reinforced Learning**
- **Task**
  - Learn how to behave successfully to achieve a goal while interacting with an external environment
  - Learn via experiences!

Reinforcement Learning



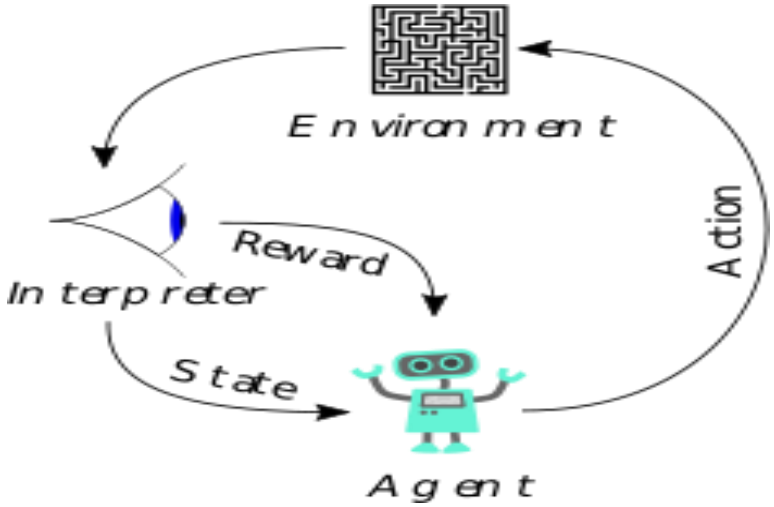
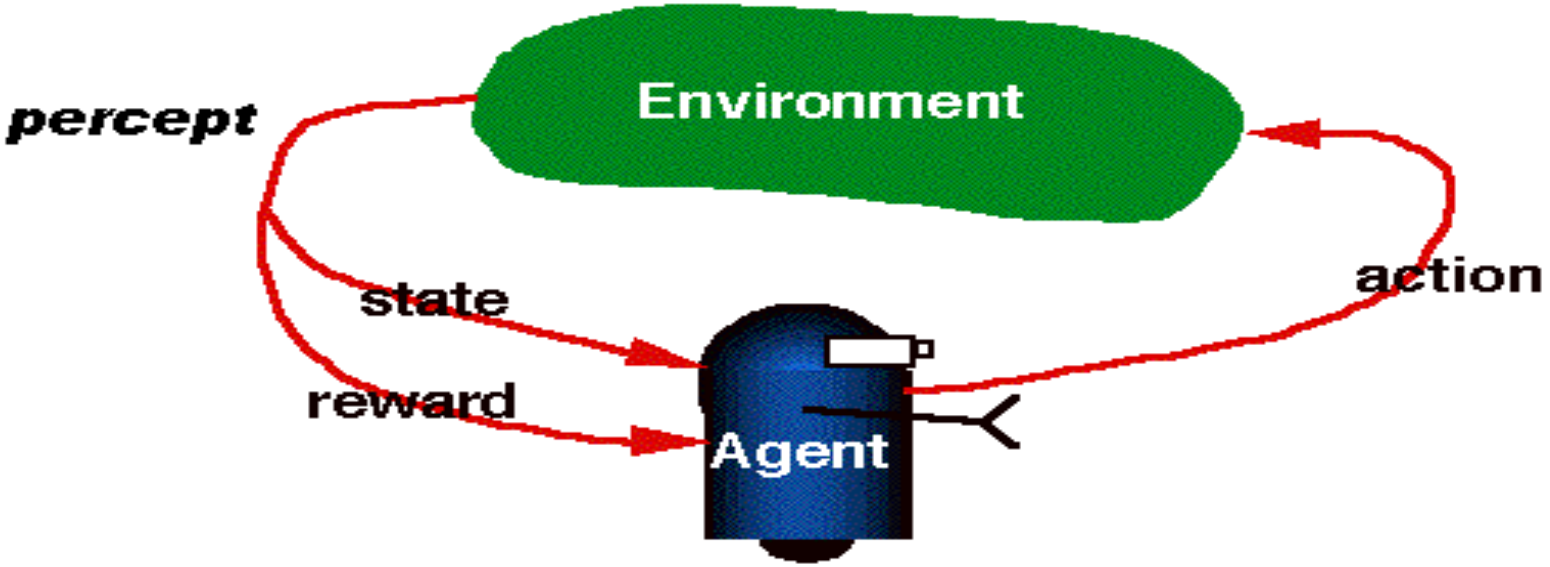


Figure: The typical framing of a Reinforcement Learning (RL) scenario: an agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent.

RL is Learning from Interaction



A Percept is the input that an intelligent agent is perceiving at any given moment.

The agent is one who takes decisions based on the rewards and punishments.

Example of a batsman in cricket. He tries to hit the ball if he misses he gets a negative point. If he hits the ball then he gets a reward.

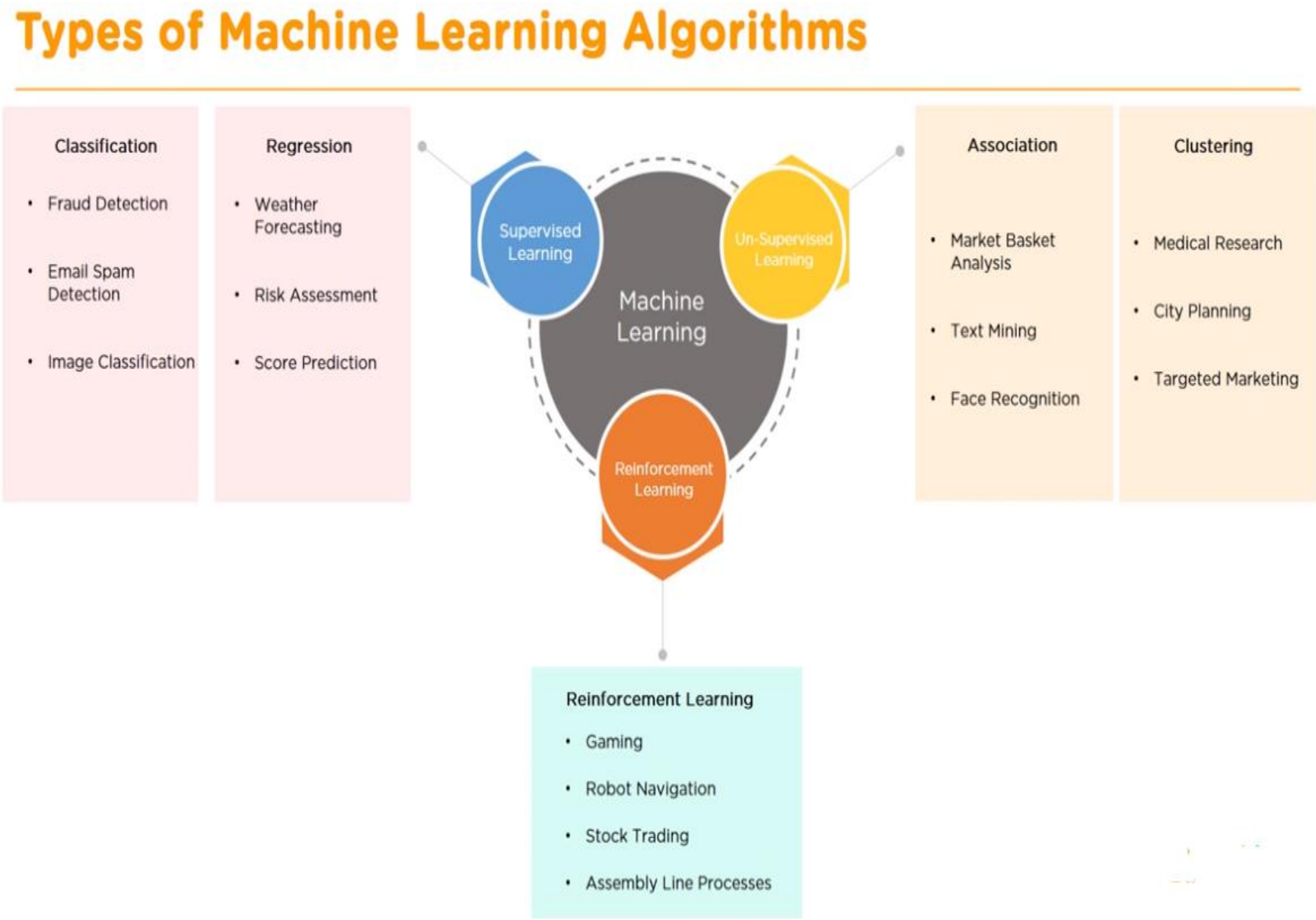
Reinforcement Learning Model

- Each percept is enough to determine the State(the state is accessible)
- The agent can decompose the Reward component from a percept.
- **The agent task:**
  - To find a optimal policy,
  - Mapping states to actions, that maximize long-run measure of the reinforcement

Supervised v/s Unsupervised v/s Reinforcement Learning

Criteria	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Learns by using labelled data	Trained using unlabelled data without any guidance.	Works on interacting with the environment
Type of data	Labelled data	Unlabelled data	No – predefined data
Type of problems	Regression and classification	Association and Clustering	Exploitation or Exploration
Supervision	Extra supervision	No supervision	No supervision
Algorithms	Linear Regression, Logistic Regression, SVM, KNN etc.	K – Means, C – Means, Apriori	Q – Learning, SARSA
Aim	Calculate outcomes	Discover underlying patterns	Learn a series of action
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self Driving Cars, Gaming, Healthcare

Types of Machine Learning Algorithms



Framework of ML model development

- **Problem or Opportunity Identification**
- **Feature Extractions- Collection of Relevant Data**

Framework of ML model development

- **Data Pre-processing-Data Cleaning and Feature Engineering**
- **Feature Engineering -**Algorithms require features with some specific characteristic to work properly.
  - **Preparing the proper input dataset, compatible with the machine learning algorithm requirements.**
  - **Improving the performance of machine learning models.**
    - **Example :** For example, if ***X*1 and *X*2** are two features that are captured in the original data.
    - We can derive new features by taking ratio **(*X*1/*X*2)** and product **(*X*1*X*2).**

Framework of ML model development

- **Model Building**
- **Communication and Deployment of the Data Analysis**-

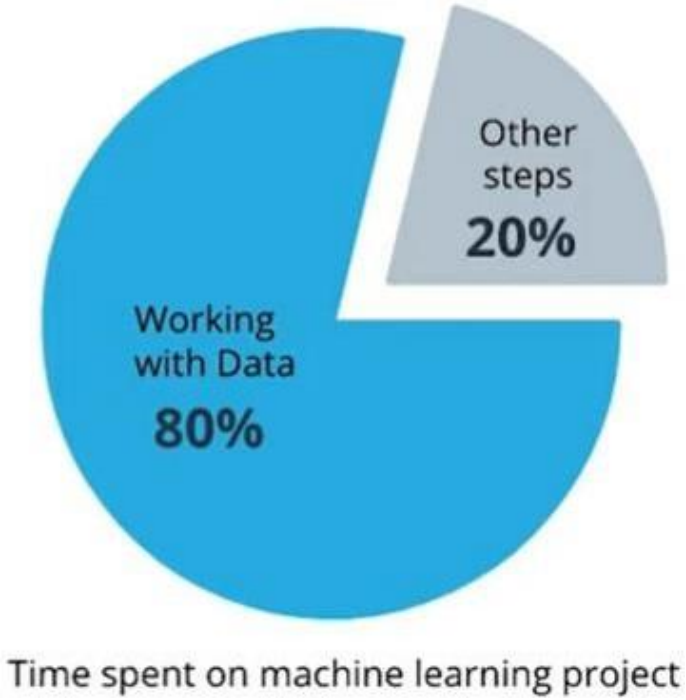
In which format you will deliver the final outcome/Product-Tabular format, Graphical format etc. Data visualization techniques are used.

Steps Involved in ML Projects

- **Build the dataset**
- **Preparing the Data**
- **Model selection/Choosing a Model**
- **Train the model**
- **Evaluate the model**
- **Hyperparameter Tuning**
- **Prediction**

Steps Involved in ML Projects

-Build the dataset



Time spent on machine learning project

Source: Forbes survey back in 2016 on 80 data scientists

Steps Involved in ML Projects

-Build the dataset

Steps Involved in ML Projects

-Build the dataset

Data Collection

- Find and collect data related to the problem you have defined.
- Supervised learning → Labeled Data
- Unsupervised learning → Unlabeled Data

Steps Involved in ML Projects

-Build the dataset

Data inspection is very important. To check data integrity, incomplete data, missing data

Data Inspection

Explore your dataset looking for

- Outliers
- Missing or incomplete data
- Transform your dataset

Many **Machine Learning algorithms** are sensitive to the **range and distribution of attribute values** in the input data.

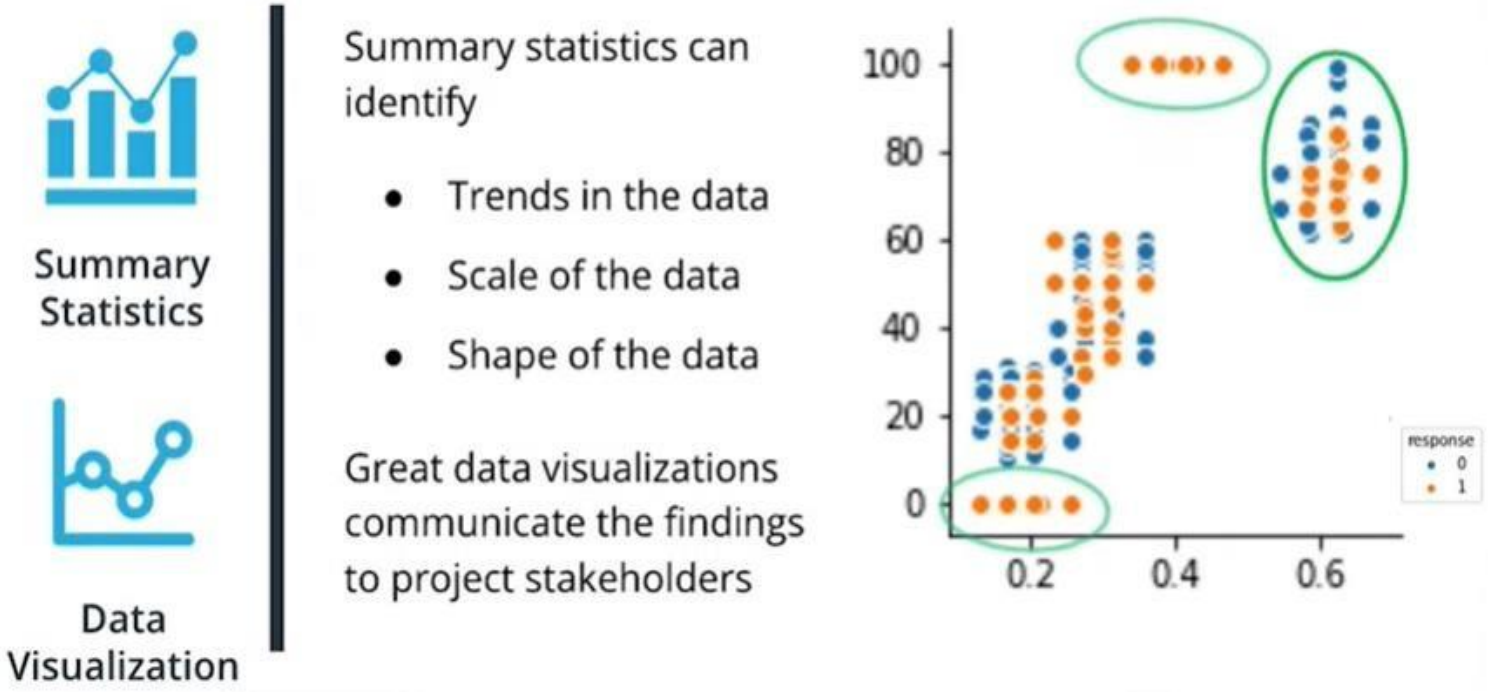
Sometimes a dataset can contain **extreme values** that are **outside the range** of what is expected and unlike the other data called as **outliers**

Outliers in input data can **skew and mislead the training process** of machine learning algorithms resulting in **longer training times, less accurate models and ultimately poorer results.**

Steps Involved in ML Projects-Build the dataset

Summary Statistics are used to summarize a set of observations, in order to communicate the largest amount of information as simply as possible.

Statistical formules like mean, Median, Standard Deviation used to get the data.



Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps eg.-**Tableau**.

Steps Involved in ML Projects-Build the dataset

Data visualization –**What is Tableau?**

Tableau is a visual analytics platform transforming the way we use data to solve problems—empowering people and organizations to make the most of their data. **Tableau helps** people and organizations be more data-driven. It helps people see and understand their data.

What is Tableau A Tableau Overview (720p).mp4

Steps Involved in ML Projects

Data pre-processing/Preparing the Data

- Dealing with missing data
- Handling Categorical Data
- Ordinal features
- Nominal features

Steps Involved in ML Projects

Data pre-processing/Preparing the Data

- Handling Categorical Data
  - Ordinal features (for example size)
    - Mapping ordinal features: L:2, M:1, S:0.
  - Nominal features (for example colour)
    - Encoding nominal class labels: Different colors-One hot encoding
  - **One-hot encoding**-It transforms features into binary form

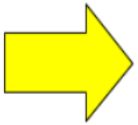
Many algorithms and also scatterplots are not compatible with non-numerical

Steps Involved in ML Projects

Data pre-processing/Preparing the Data

- Handling Categorical Data
  - Nominal features
    - Encoding nominal class labels: Different Colors-One hot Encoding

Color
Red
Red
Yellow
Green
Yellow



Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1

Steps Involved in ML Projects

Data pre-processing/Preparing the Data

- Redundant Data
- Structured Data-Data that is collected from a known method and can be neatly arranged
- Unstructured Data-Data that has no predefined format or organization, making analysis more difficult
- **Feature Scaling-** It is used to bring all values into the range [0,1].
  - **Normalization** →

$$X_{changed} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

It includes eliminating **unstructured data** and **redundancy (duplicates)** in order to ensure logical data storage.

  - **Standardization** →

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation

It is the **process of bringing data into a uniform format** that allows analysts and others to research, analyze, and utilize the data.

Steps Involved in ML Projects

❑ Selecting Meaningful Features

Overfitting –

- A statistical model is said to be overfitted when we train it with a lot of data *(just like fitting ourselves in oversized pants!)*.
- When a model gets trained with so much data, it starts learning from the noise and inaccurate data entries in our data set.
- Then the model does not categorize the data correctly, because of too many details and noise

Solution-Reduce data's dimensionality, No. of features, Reduce model complexity etc.

Underfitting-

- A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data.
- It destroys the accuracy of our machine learning model.
- It usually happens when we have less data to build an accurate model.

Solution-Increase model complexity, Increase the number of features, remove noise from the data, increase the number of epochs etc.

Steps Involved in ML Projects

- **Splitting Data Into Subsets**
  - Training
  - Testing
  - Validation

Training Dataset-The sample of data used to fit the model.

The actual dataset that we use to train the model (weights and biases in the case of a Neural Network). The model sees and learns from this data.

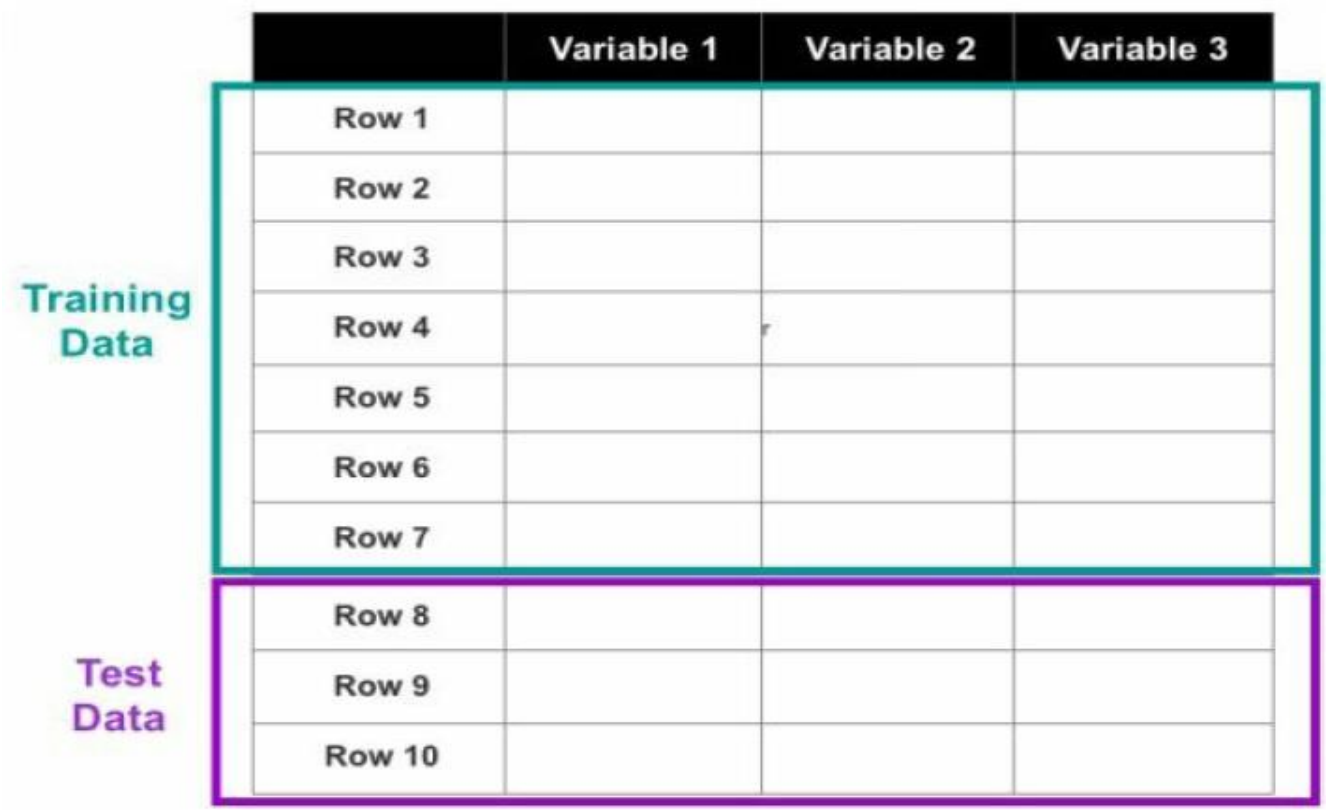
Testing Dataset-The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. The Test dataset provides the gold standard used to evaluate the model. It is only used once a model is completely trained.

Validation Dataset-The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model **hyperparameters.**

Data validation means checking the accuracy and quality of source data before using, importing or otherwise processing data.

Training and Test Data

- Once you have **cleaned** your dataset, **data is split into training data and test data.**
- It is very important not to test your model with the same data that you used for training.
- The ratio of the two splits should be approximately 70/30 or 80/20.
- Before you split your data, it is important that you randomize all rows in the dataset.



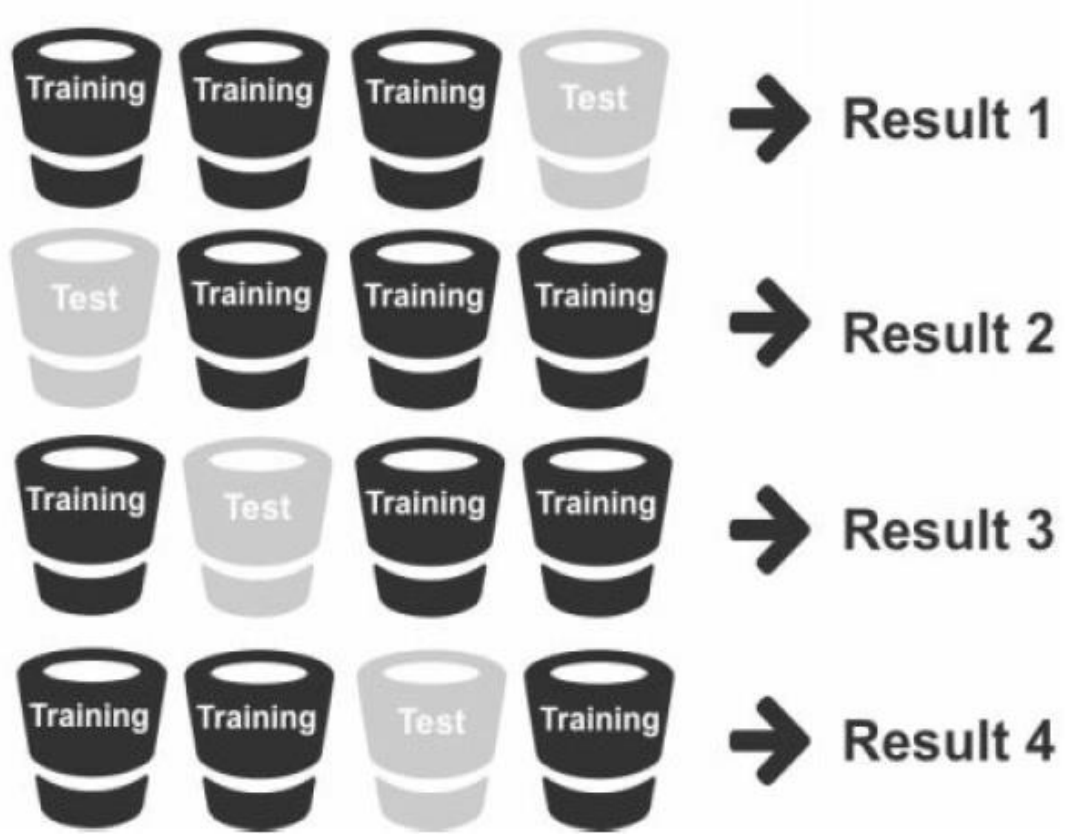
	Variable 1	Variable 2	Variable 3
Row 1			
Row 2			
Row 3			
Row 4			
Row 5			
Row 6			
Row 7			
Row 8			
Row 9			
Row 10			

Training and Test Data

- Rather than splitting the data into two segments (one for training and one for testing), we can implement what is known as **cross validation.**
- **Cross validation** maximizes the availability of training data by splitting data into various combinations and testing each specific combination.
- Two Methods-**Exhaustive Cross Validation** & **Non-exhaustive cross validation**, known as **k-fold validation**.
- **Exhaustive cross validation** involves finding and testing all possible combinations to divide the original sample into a training set and a test set.
- **The k-fold validation** technique involves splitting data into k assigned buckets and reserving one of those buckets to test the training model at each round.

Training and Test Data

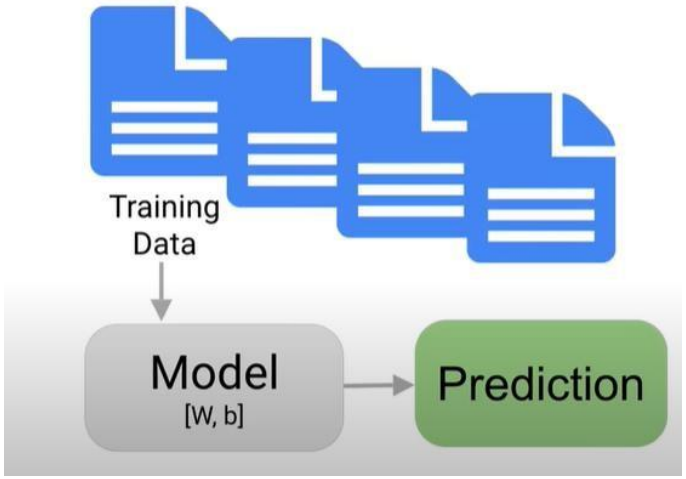
- **K-fold Validation**



- To perform k-fold validation, data are first randomly assigned to k number of equal sized buckets.
- One bucket is then reserved as the test bucket and is used to measure and evaluate the performance of the remaining (k-1) buckets.

Steps Involved in ML Projects

- Why not having only two sets, training and testing?
- Answer is:
- **Hyperparameters tuning**
- Hyperparameters are settings on the model which are not changed during training but can affect how quickly or how reliably the model trains,
- E.g., the number of clusters the model should identify.



Parameters and Hyperparameters

What are Model Parameters

- Variables internal to the neural network.
- Value can be estimated right from the data.

Model Parameters

- Required by the model to make predictions
- Values define the skill of the model
- **Estimated directly from data**
- **Not set manually**
- Saved as part of the learned model
- Examples: Weights, Biases etc.

What are Model Hyperparameters?

Configurations external to the neural network.

Value cannot be estimated right from the data.

Model Hyperparameters

No clear-cut way to find the best value.

When a DL algorithm is tunes, you are really tuning the Hyperparamters.

If you have to manually specify a parameter, it is a hyperparamter.

Example:-Learning rate, C in SVMs

Summary

Summary

Model Parameters ⟶ Estimated from the data

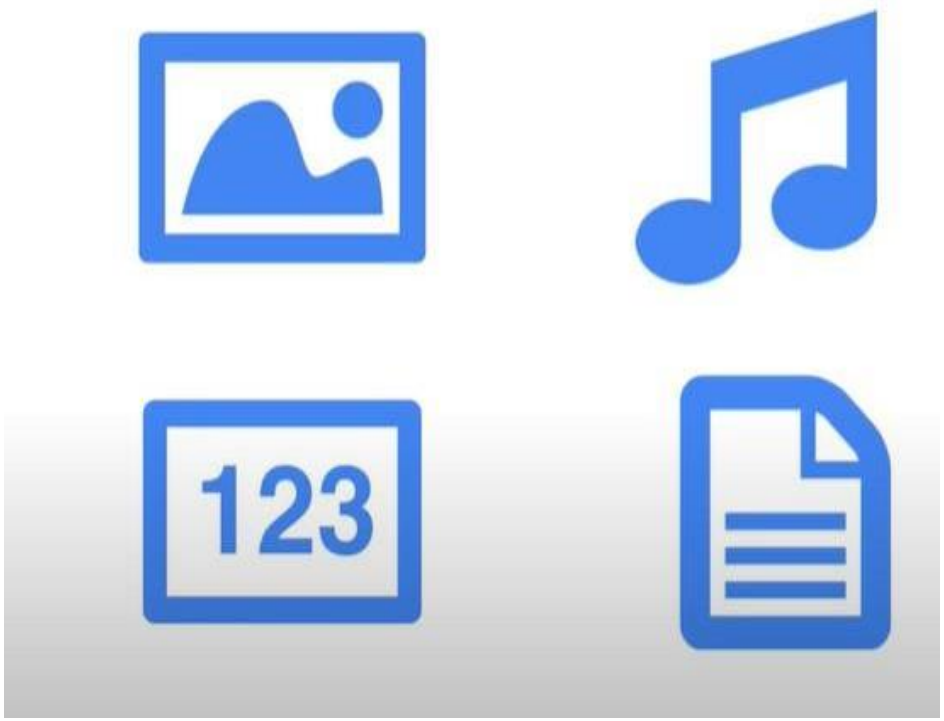
Model Hyperparameters ⟶ Can’t be estimated from the data.

Model hyperparameters are often referred to as parameters because they are the parts of the machine learning that must be set manually and tuned.

Steps Involved in ML Projects

Model selection/Choosing a Model

- Different algorithms are for different tasks; choose the right one



Steps Involved in ML Projects

Train the model

What does a model training algorithm actually do?

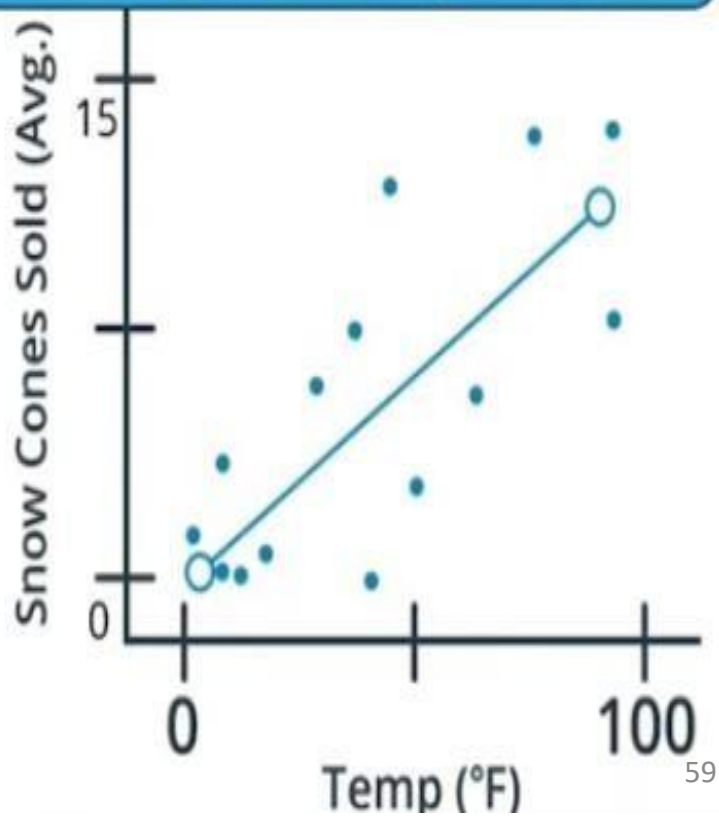
Iteratively update ***model parameters*** to minimize some ***loss function***.

- **Model Parameters**

Configuration that changes how the model behaves

- ***Loss Function***

Measurement of how close the model is to its goal



Steps Involved in ML Projects

Train the model

What does a model training algorithm actually do?

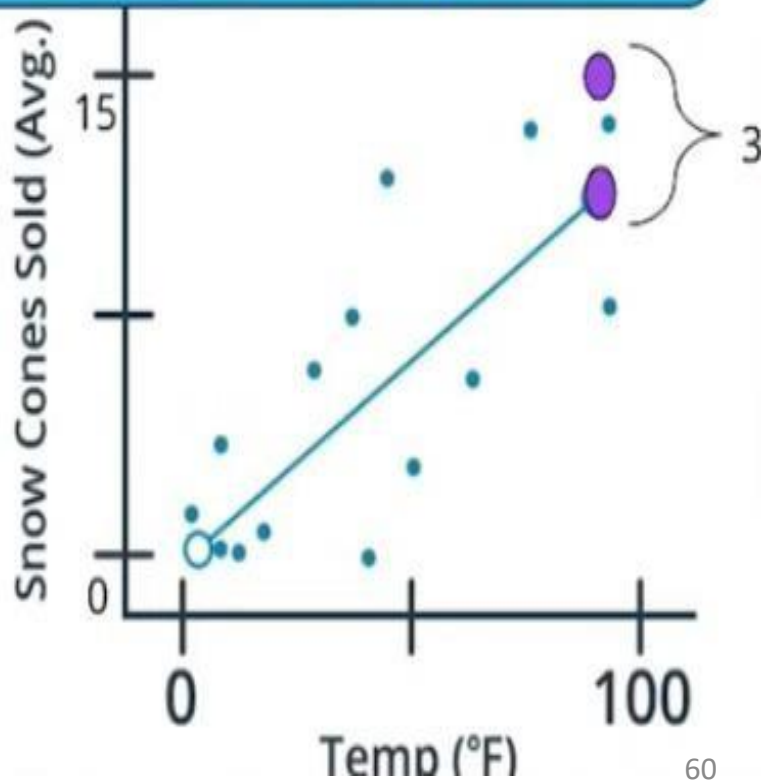
Iteratively update ***model parameters*** to minimize some ***loss function***.

- **Model Parameters**

Configuration that changes how the model behaves

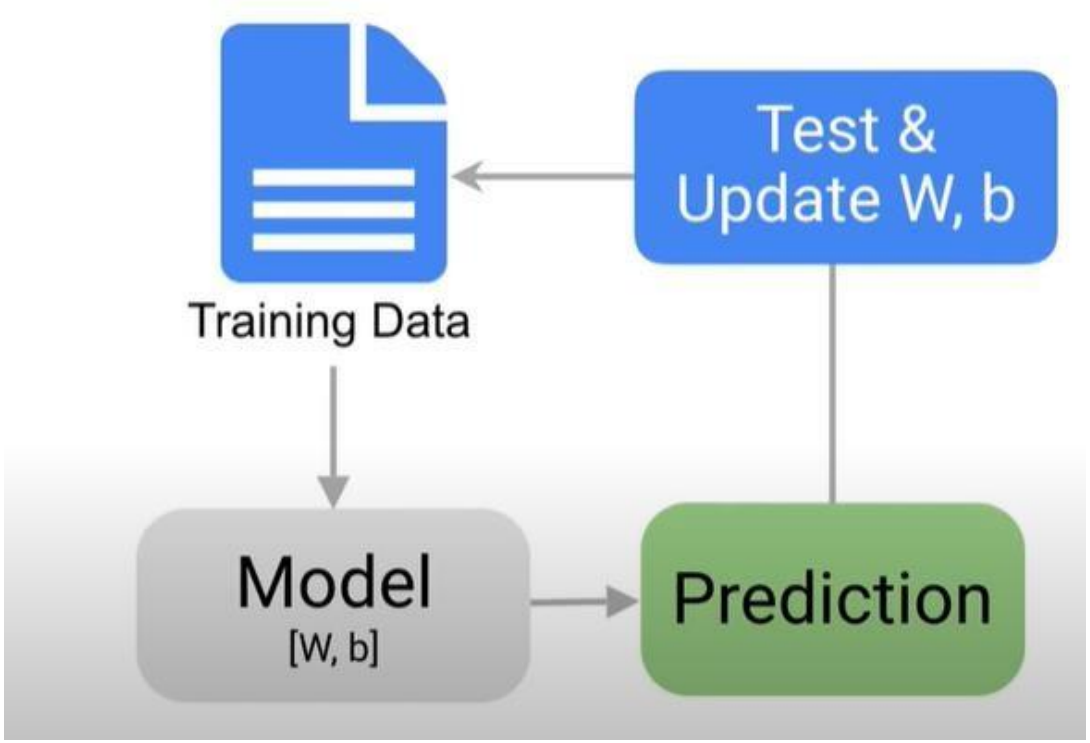
- ***Loss Function***

Measurement of how close the model is to its goal



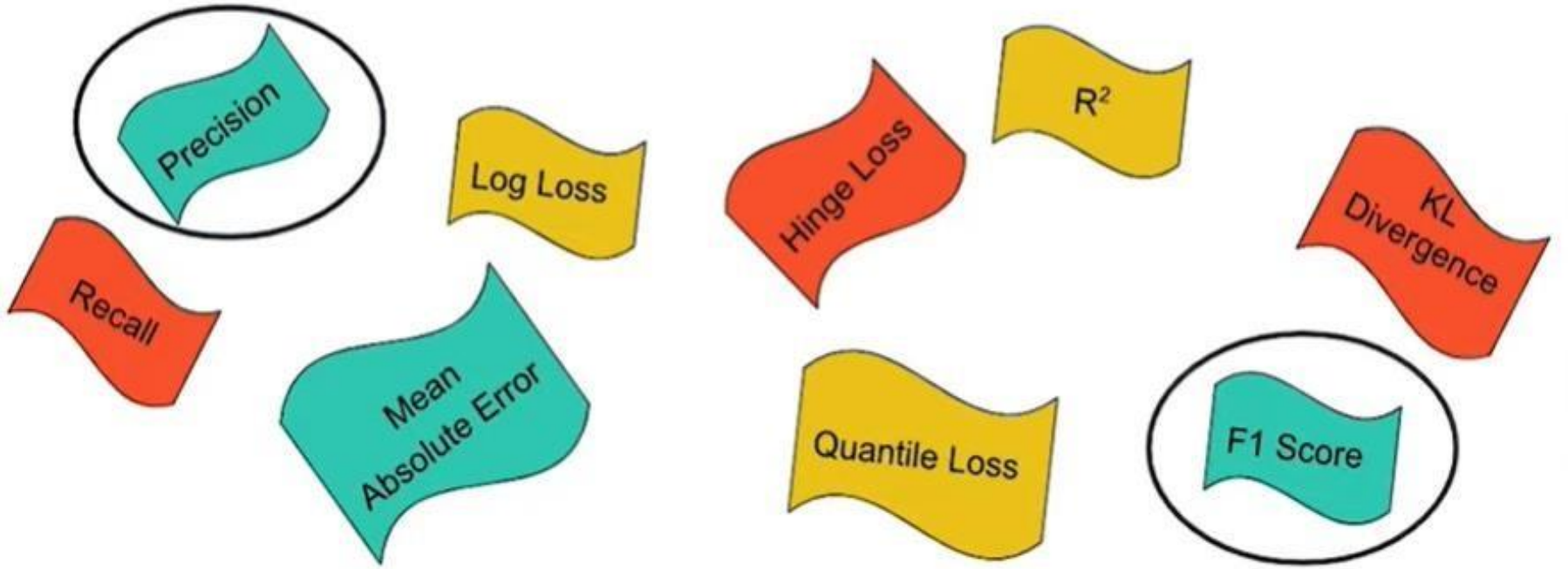
Steps Involved in ML Projects

Train the model



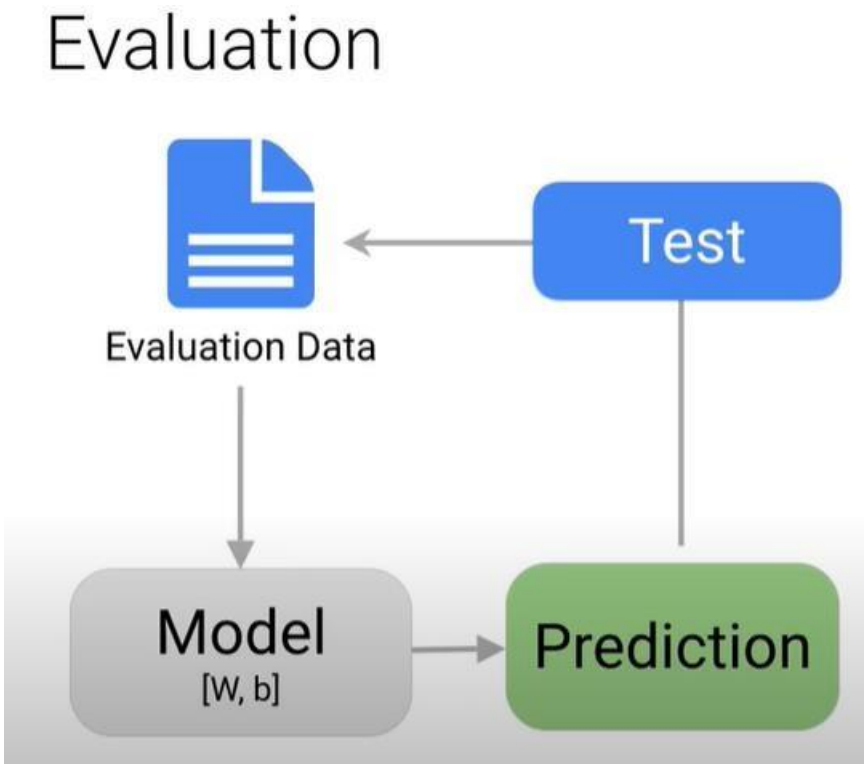
Steps Involved in ML Projects

Evaluate the Trained model



Steps Involved in ML Projects

Evaluate the model




ThankYou
Merci
Salamat
Kop
Tack
Hvala
Xie
kun
Gra
Dankie
Kruthagnathalu
Arigatou
Dhonnobaad
Asante
Hain
Dhan
daa
Shukriya
Khopjai
Dhanyavad
Gomapsupnida
Kun
Euxaristo
Danke
dank
Terima
Enkosi
Fyrir
Blagodariya
Gracias
Nandree
Grazzi
raibh
krap
Dakujem
Daw
Waad
Dhanyavadaalu
Dank
Takk
Gamsahapnida
Te°ekkür
Dekuju/Dekujeme
Diolch
Kiitos
Sheun
umesc
Shnorhakalutiun
Kasih
Mamnoon
Todah
Shokriya
Ngiyabonga
Dziękuję
Shokrun
Spaas
Mul
Ačі
Cam
or
al
Dhanyavaad
Go
Grazie
Falemiderit
Or
ederim