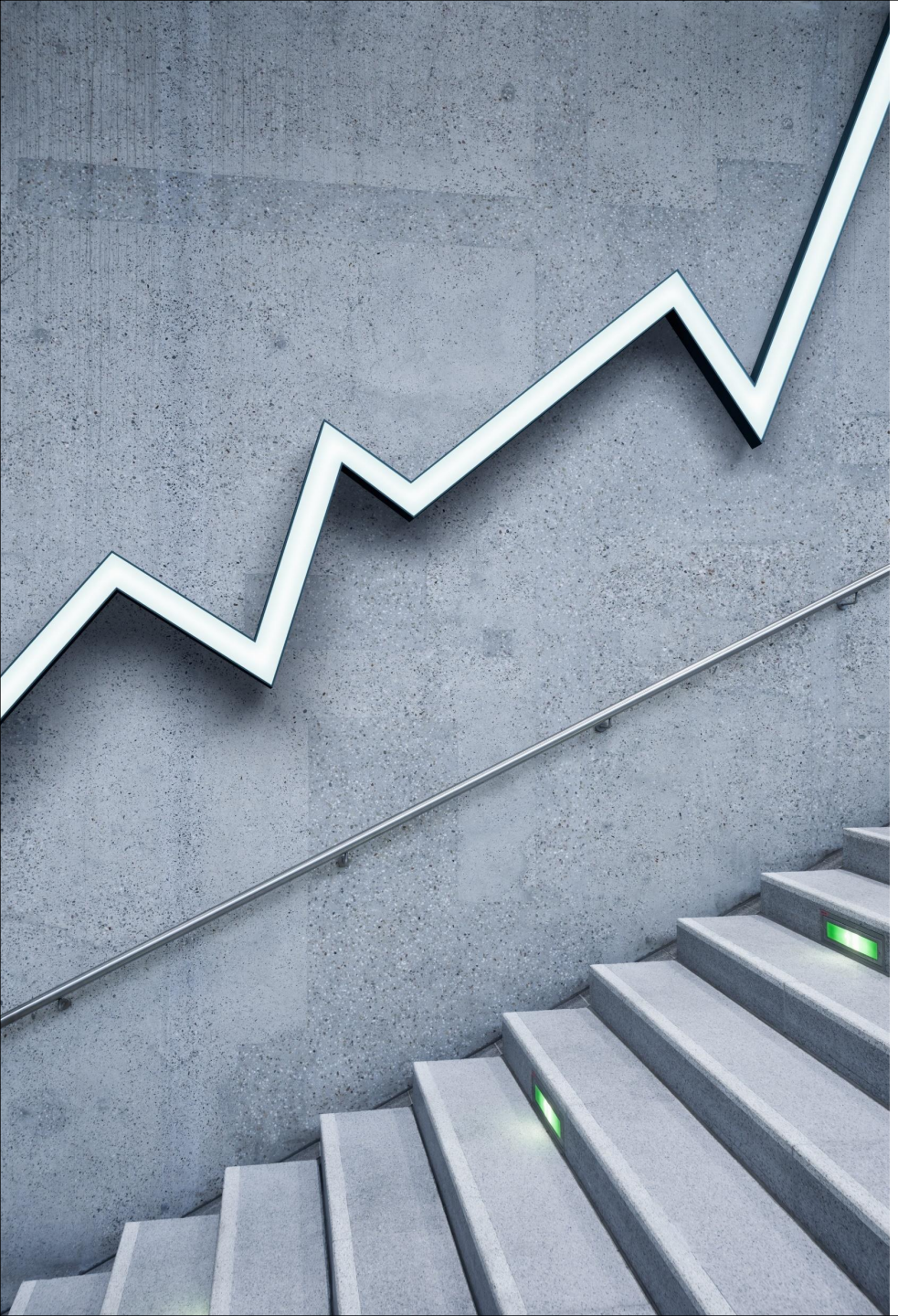
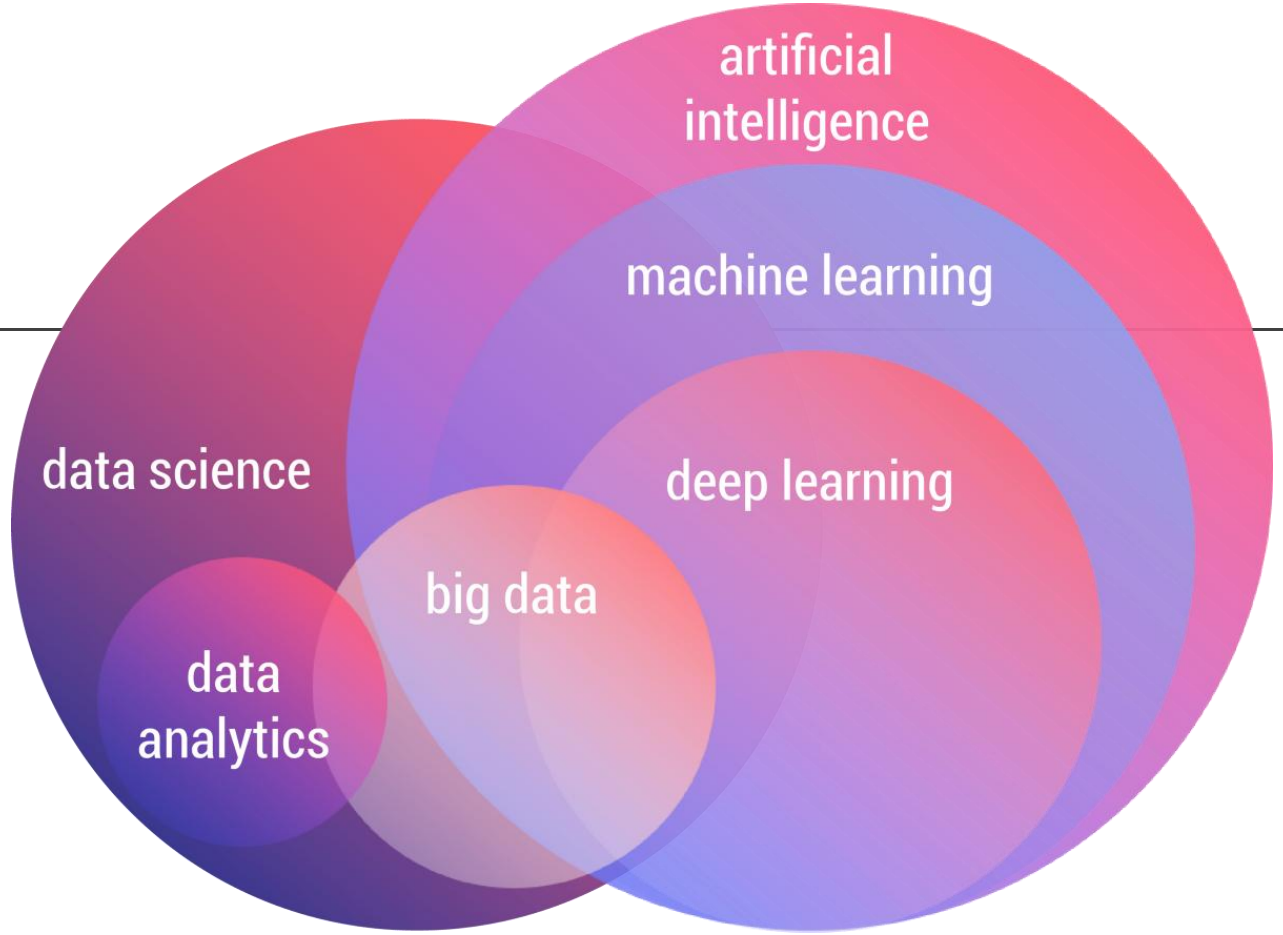


Machine learning

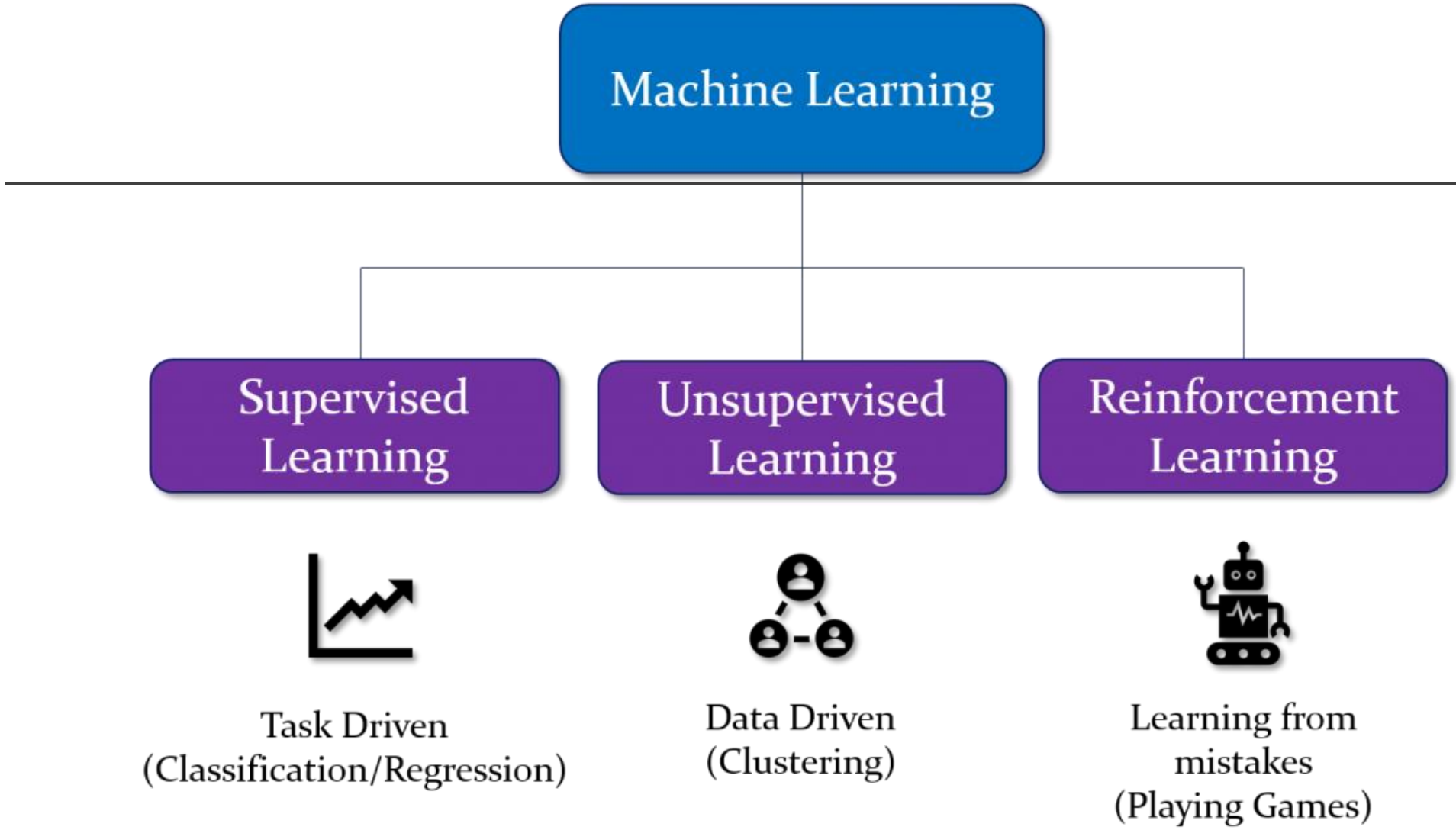

artificial
intelligence
machine learning
data science
deep learning
big data
data
analytics

ML

Machine learning

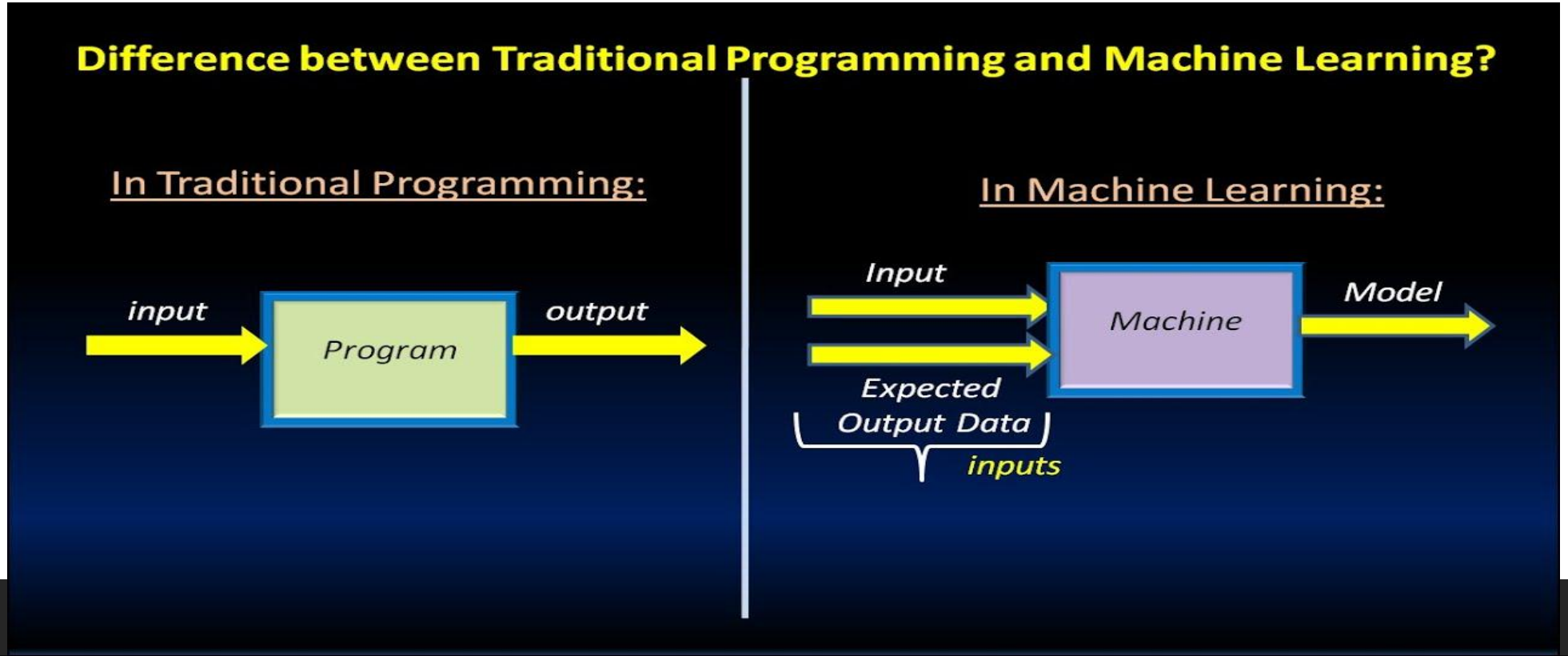
Machine learning uses artificial intelligence to enable machines to learn and predict outcomes more accurately without being explicitly programmed to do so. This makes them more similar to humans. The machine learning models continuously learns and improves its performance with the use of necessary data.

Types of Machine Learning



TRADITIONAL PROGRAMMING

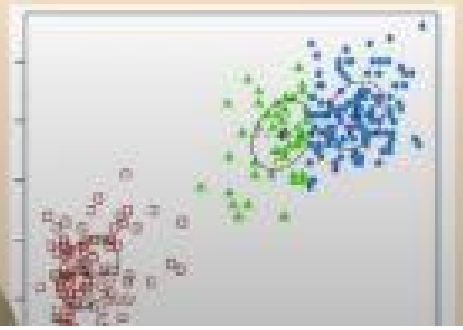
MACHINE LEARNING



Unsupervised Learning

- Unlabeled data X
- Learn X
- Generate fakes, insights

"This product does what it is supposed to. I always keep three of these in my kitchen just in case ever I need a replacement cord."



Supervised Learning

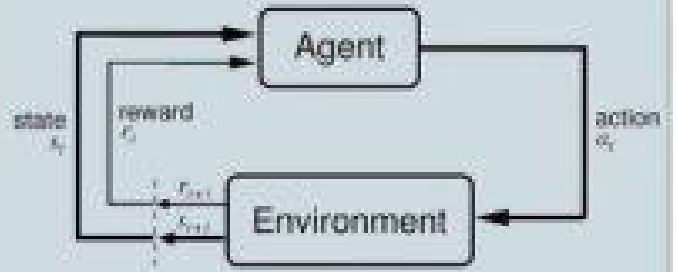
- Labeled data X and Y
- Learn X -> Y
- Make Predictions

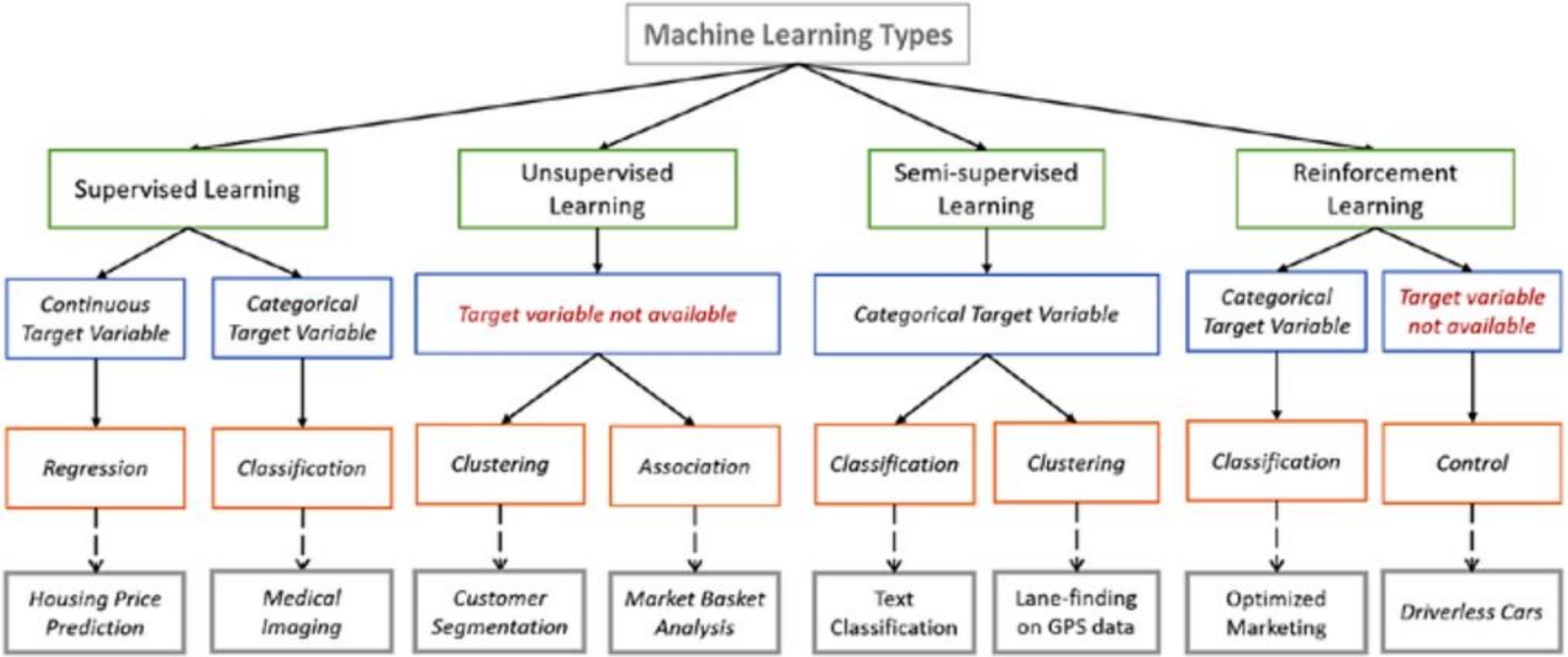
→ *cat*

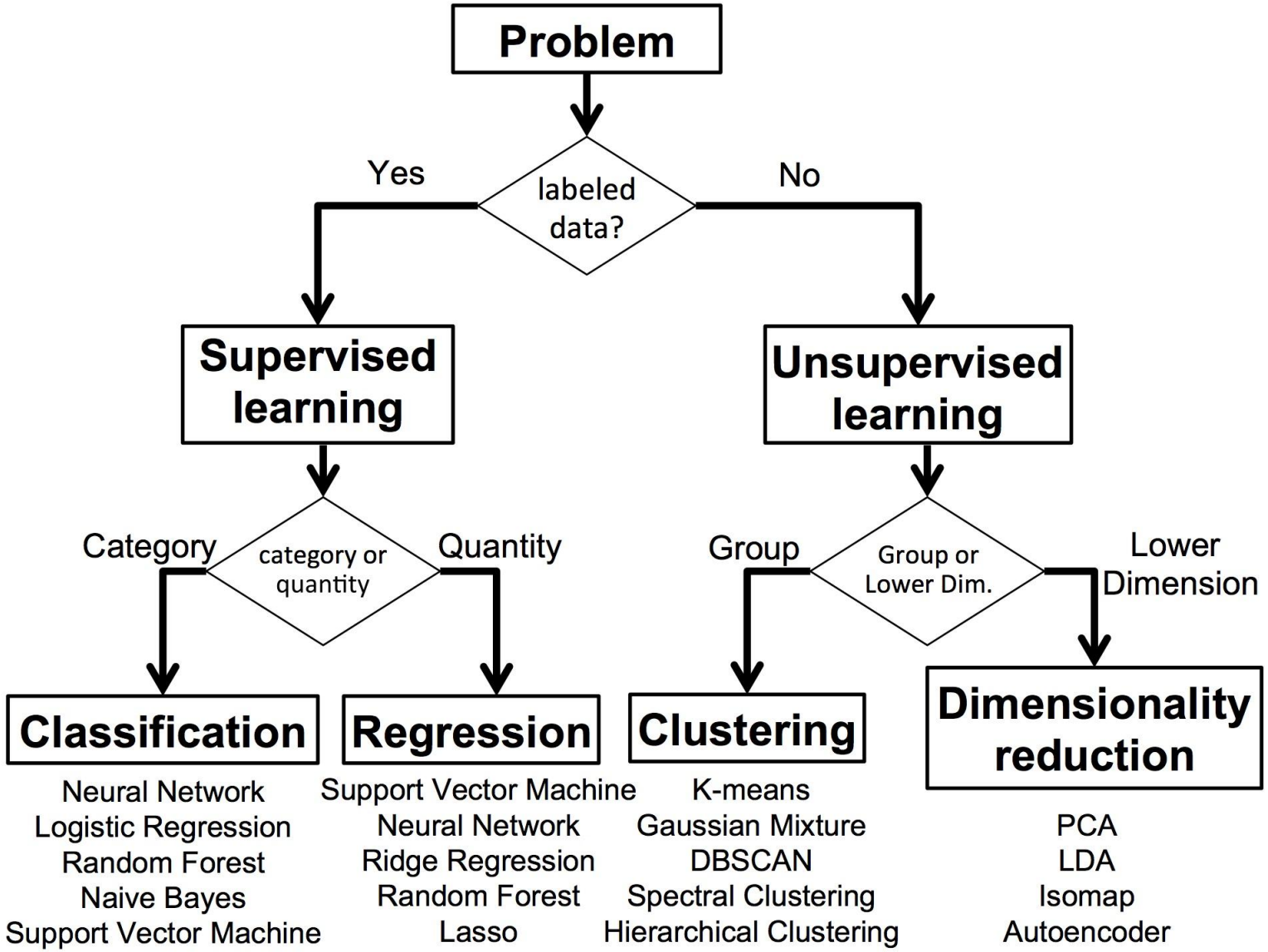
→ *"Hey Siri"*

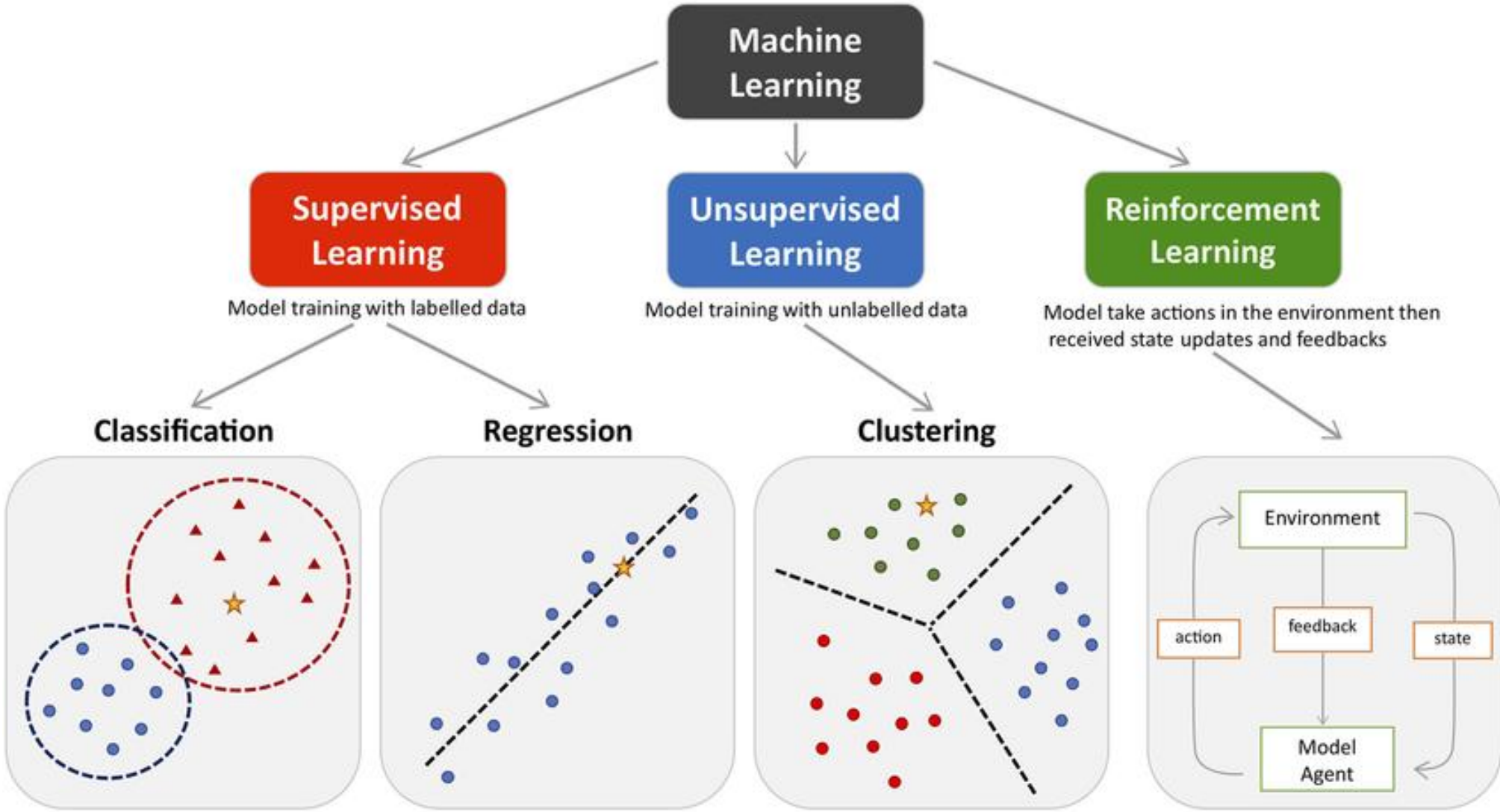
Reinforcement Learning

- Learn how to take Actions in an Environment

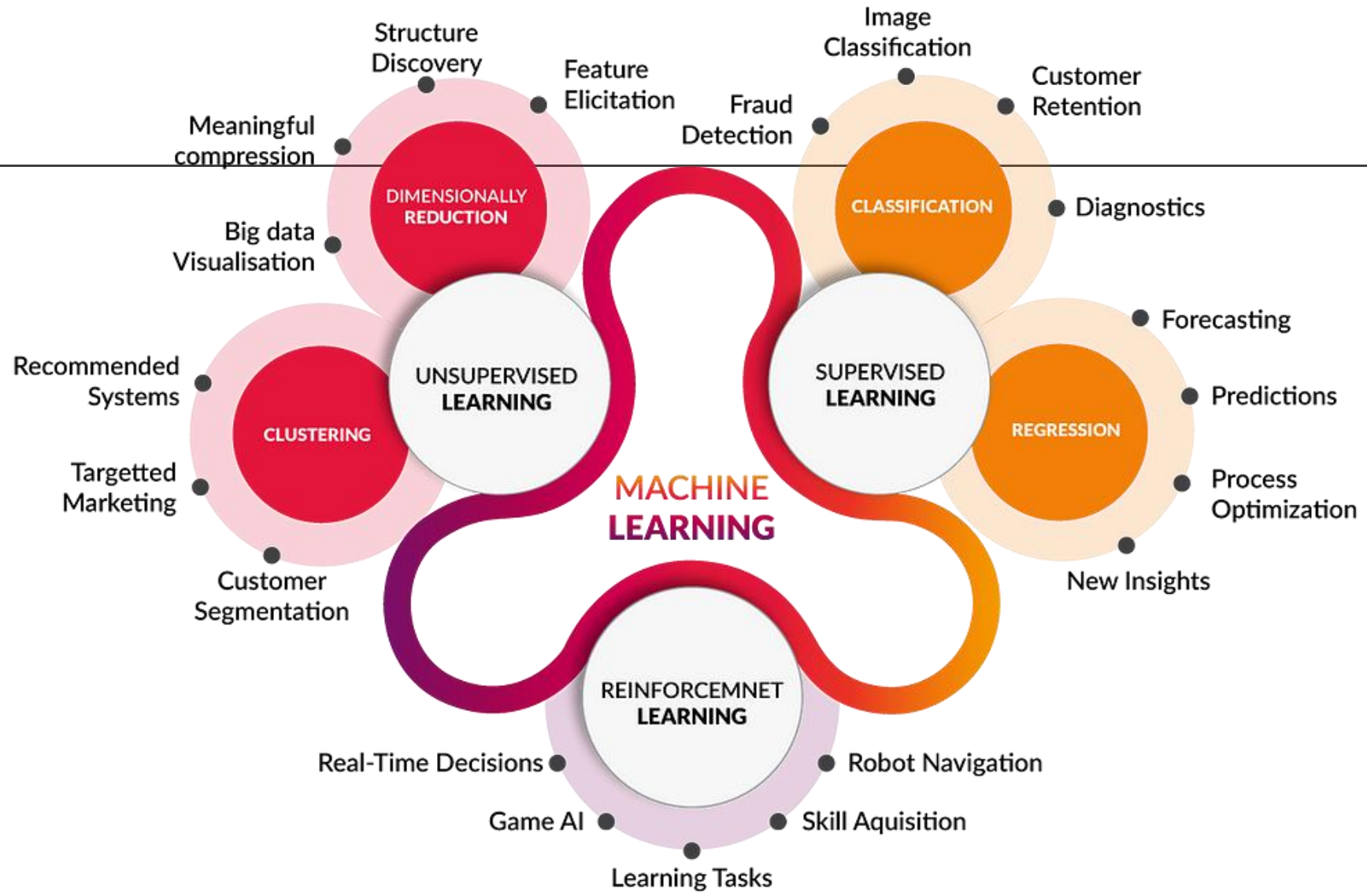



Machine Learning Types
Supervised Learning
Unsupervised Learning
Semi-supervised Learning
Reinforcement Learning
Continuous Target Variable
Categorical Target Variable
Target variable not available
Categorical Target Variable
Categorical Target Variable
Target variable not available
Regression
Classification
Clustering
Association
Classification
Clustering
Classification
Control
Housing Price Prediction
Medical Imaging
Customer Segmentation
Market Basket Analysis
Text Classification
Lane-finding on GPS data
Optimized Marketing
Driverless Cars


Problem
labeled data?
Yes
No
Supervised learning
Unsupervised learning
category or quantity
Category
Quantity
Group or Lower Dim.
Group
Lower Dimension
Classification
Regression
Clustering
Dimensionality reduction
Neural Network
Logistic Regression
Random Forest
Naive Bayes
Support Vector Machine
Support Vector Machine
Neural Network
Ridge Regression
Random Forest
Lasso
K-means
Gaussian Mixture
DBSCAN
Spectral Clustering
Hierarchical Clustering
PCA
LDA
Isomap
Autoencoder


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Model training with labelled data
Model training with unlabelled data
Model take actions in the environment then received state updates and feedbacks
Classification
Regression
Clustering
Environment
action
feedback
state
Model Agent

Examples use case



Supervised

- In Supervised learning, you train the machine using data which is well **"labeled."**
- It means data is already tagged with the correct answer.
- It can be compared to learning which takes place in the presence of a supervisor or a teacher.
- A supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data.
- One disadvantage of this learning method is that the dataset has to be hand-labeled either by a Machine Learning Engineer or a Data Scientist. This is a very *costly process*, especially when dealing with large volumes of data.

Unsupervised

- Unsupervised Learning is a machine learning technique in which the users do not need to supervise the model.
- Instead, it allows the model to work on its own to discover patterns and information that was previously undetected.
- It mainly deals with the unlabeled data.
- Unsupervised learning problems are grouped into clustering and association problems.
- The most basic disadvantage of any Unsupervised Learning is that it's application spectrum is limited.

Semi-supervised

- Semi-supervised learning is the type of machine learning that uses a combination of a small amount of labeled data and a large amount of unlabeled data to train models.

Intuitively, one may imagine the three types of learning algorithms as :

- Supervised learning where a student is under the supervision of a teacher at both home and school.
- Unsupervised learning where a student has to figure out a concept himself.
- Semi-Supervised learning where a teacher teaches a few concepts in class and gives questions as homework which are based on similar concepts.

Reinforcement

- Reinforcement learning is the training of machine learning models to make a sequence of decisions.
- In this approach, machine learning models are trained to make a series of decisions based on the rewards and feedback they receive for their actions.
- The machine learns to achieve a goal in complex and uncertain situations and is rewarded each time it achieves it during the learning period.
- Reinforcement learning is different from supervised learning in the sense that there is no answer available, so the reinforcement agent decides the steps to perform a task.
- The machine learns from its own experiences when there is no training data set present.

Based On	Supervised machine learning	Unsupervised machine learning
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data which is not labelled
Computational Complexity	Supervised learning is a simpler method.	Unsupervised learning is computationally complex
Accuracy	Highly accurate and trustworthy method.	Less accurate and trustworthy method.

The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest.

-- Kevin P. Murphy

ML key terms that you must know!

Features

- Features are the fields used as input.
- A feature is one column of the data in your input set.
- For instance, if you're trying to predict the type of pet someone will choose, your input features might include age, home region, family income, etc.
- Feature means property of your training data.
- A feature is the input you have fed to the model or system.
- The value of x variable in simple linear regression are the features.

Label

- The output you get from your model after training is called a label.
- A label is the thing we're predicting.
- For example the value of y variable in simple linear regression model is the label.
- Suppose you give your model data like a person's age, height, hair length and then your model predicts whether the person is male or female. Then male or female is called the label.

Tabular data to predict whether car is stolen or not?

Features

Label

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

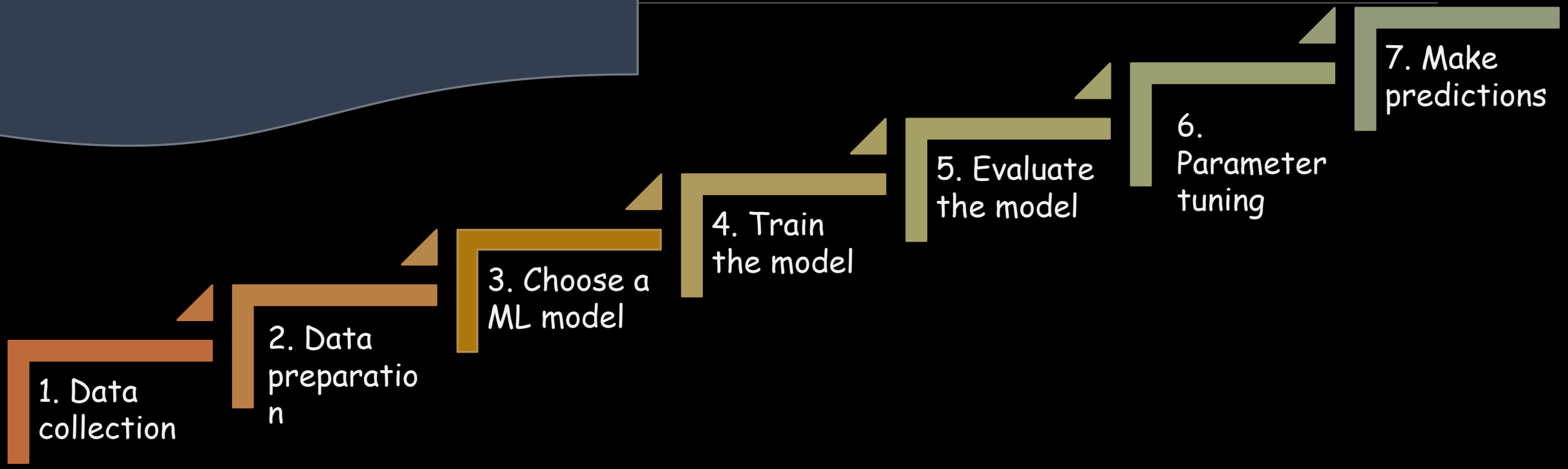
Model

- A model is the relationship between features and the label.
- An ML model is a mathematical model that generates predictions by finding patterns in your data.
- ML Models generate predictions using the patterns extracted from the input data.
- A model represents what was learned by a machine learning algorithm.
- The model is the "thing" that is saved after running a machine learning algorithm on training data and represents the rules, numbers, and any other algorithm-specific data structures required to make predictions.

https://medium.com/analytics-vidhya/parametric-and-nonparametric-models-in-machine-learning-a9f63999e233


Data
ML
Algorithm
ML Model

Steps in ML


1. Data collection
2. Data preparation
3. Choose a ML model
4. Train the model
5. Evaluate the model
6. Parameter tuning
7. Make predictions

1. Data collection

Data collection is the process of gathering and measuring information from countless different sources.

This is a critical first step that involves gathering data from various sources such as databases, files, and external repositories.

Before starting the data collection process, it's important to articulate the problem you want to solve with an ML model.

2. Data Preparation

Data preparation/pre-processing techniques generally refer to the addition, deletion, or transformation of training set data.

Since the collected data may be in an undesired format, unorganized, or extremely large, further steps are needed to enhance its quality. The three common steps for preprocessing data are formatting, cleaning, and sampling.

Data preparation (also referred to as "data preprocessing") is the process of transforming raw data so that data scientists and analysts can run it through machine learning algorithms to uncover insights or make predictions.

3. Choose a ML model

For different purpose, different ML models are available. So it depends on the need that which ML model must be selected.

The choice of ML model to be selected depends on many factors like the problem statement and the kind of output you want, type and size of the data, the available computational time, number of features, and observations in the data, etc.

4. Train the model

The process of training an ML model involves providing an ML algorithm (that is, the *learning algorithm*) with training data to learn from.

Let's say that you want to train an ML model to predict if an email is spam or not spam.
You would provide ML model with training data that contains emails for which you know the target (that is, a label that tells whether an email is spam or not spam). Then the model should be trained by using this data, resulting in a model that attempts to predict whether new email will be spam or not spam.

5. Evaluate the model

Model evaluation is a method of assessing the correctness of models on test data. The test data consists of data points that have not been seen by the model before.

There are two methods of evaluating models in data science, Hold-Out and Cross-Validation.
To avoid overfitting, both methods use a test set (not seen by the model) to evaluate model performance.

6. Parameter Tuning

Each model has its own sets of parameters that need to be tuned to get optimal output.

For every model, our goal is to minimize the error or say to have predictions as close as possible to actual values. This is one of the cores or say the major objective of hyperparameter tuning.

There are following three approaches to Hyperparameter tuning:

- Manual Search
- Random Search
- Grid Search

7. Make predictions

"Prediction" refers to the output of an algorithm after it has been trained on a historical dataset.

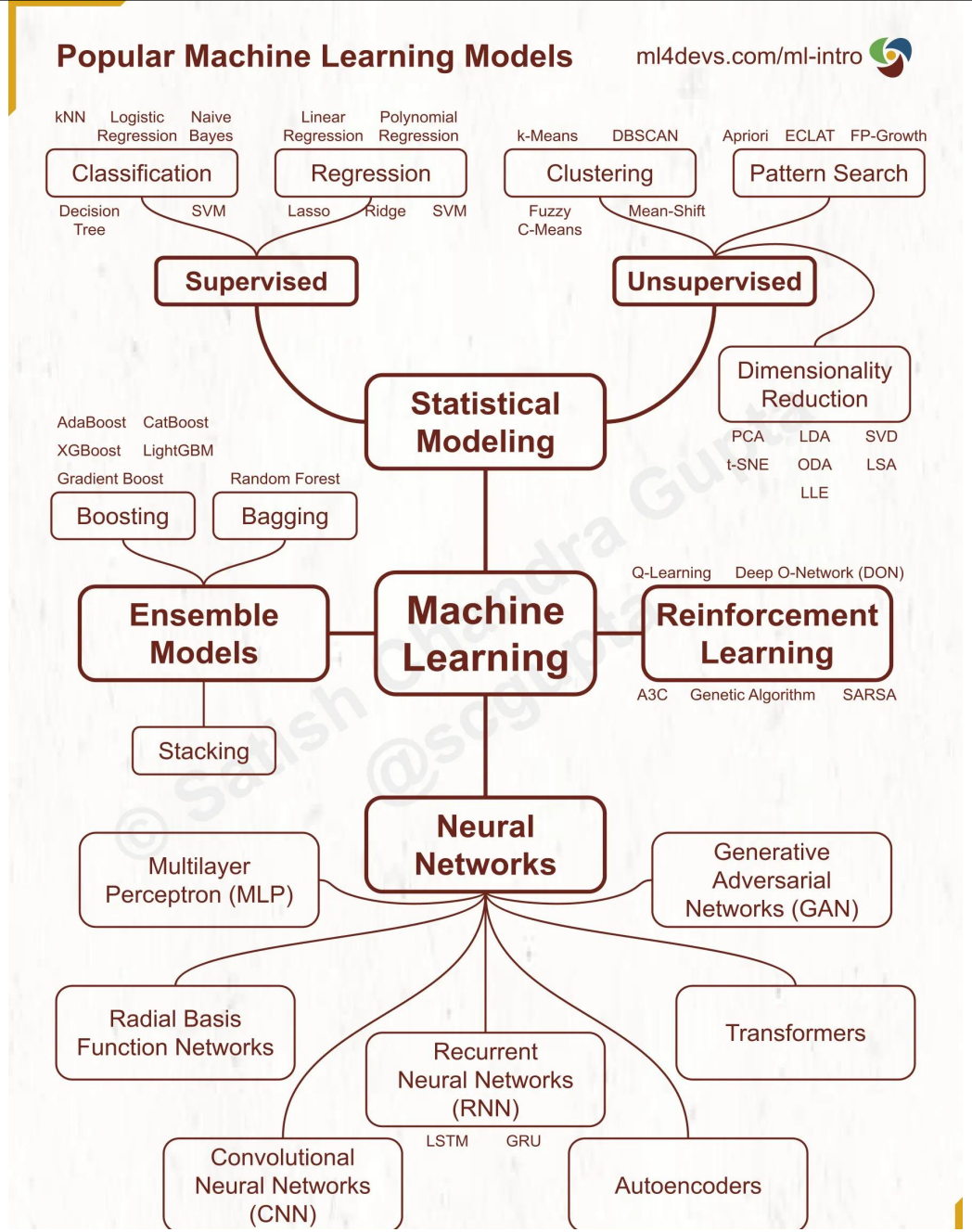
Machine learning has two main goals:
prediction and *inference*.

After you have a model, you can use that model to generate predictions which means to give your model the inputs it has never seen before and obtain the answer the model has predicted.

In addition to making predictions on new data, you can use machine-learning models to better understand the relationships between the input features and the output target which is known as inference.

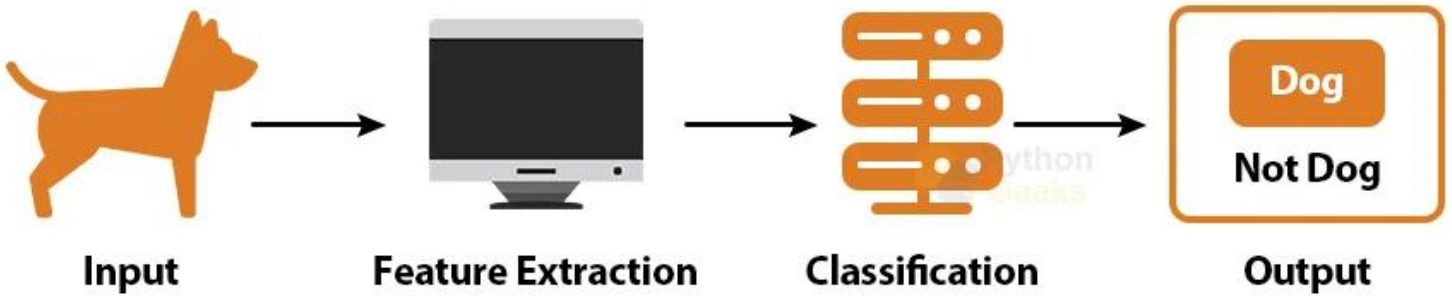
Applications Of ML

- Traffic Alerts
- Social Media
- Transportation and Commuting
- Products Recommendations
- Virtual Personal Assistants
- Self Driving Cars
- Dynamic Pricing
- Google Translate
- Online Video Streaming
- Fraud Detection



https://pythonnumericalmethods.studentorg.berkeley.edu/notebooks/chapter25.01-Concept-of-Machine-Learning.html

Traditional Machine Learning



Deep Learning

Termonolgies in ML

Data lickage
Train test split
Overfitting, underfitting ,genraliszed model
Bias , variance,
Bias variance tradeoff
Hyper parameter tuning
Model trainig
Models
Cost funtions/loss functions
Error funtions
Features
Predictions
Target variable
Label/unlabel data
Assumptions of Linear regression
Feature Selection
Multicollinearity

Imbalance , balance data

Correlation
Evaluation Metrics
learning rate
y=βO +β1 X
best Fit Line
dependent or target variable or y or label data
independent variable or X or features or predictor of Y

^yi=θ1+θ2xi
θ1: intercept
θ2: coefficient of x

Gradient Descent
A gradient is nothing but a derivative that defines the effects on outputs of the function with a little bit of variation in inputs.

- *https://www.geeksforgeeks.org/ml-linear-regression/*
- *https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/*

supervised Learning
- Regressions
  - Linear Regression
- Classification
  - Logistic Regression

Introduction to ML
ML Modeling Flow
Prametric and Non Parametric Algorithm
Types of ML

EDA

- EDA-Info,shape
- Handling Missing Values
- Handling Outliers
- Handling Skewness
- Data Encoding
- Feature Scaling - Normalization and Standardization
- Feature Engineering

https://www.geeksforgeeks.org/difference-between-parametric-and-non-parametric-methods/