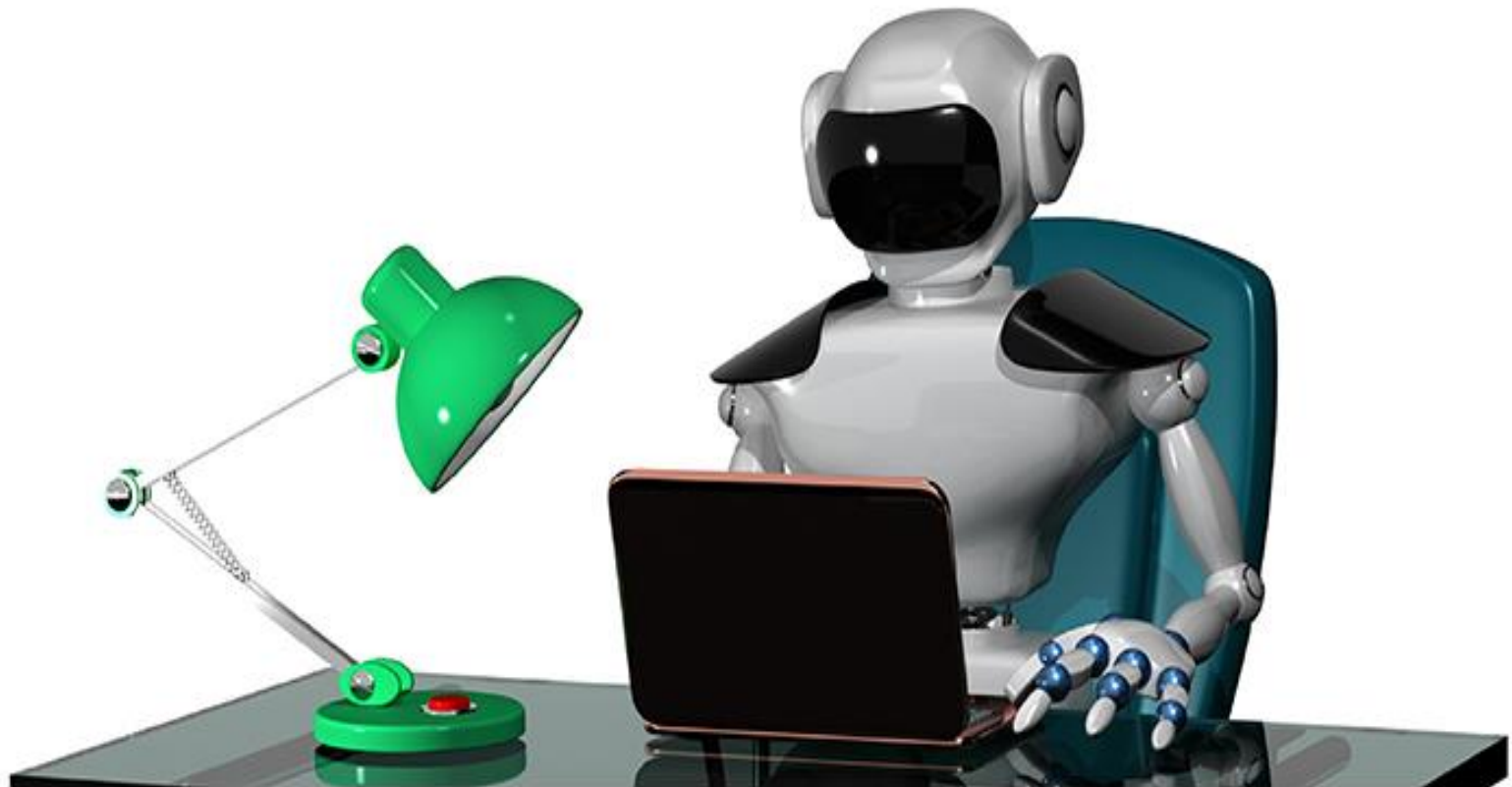
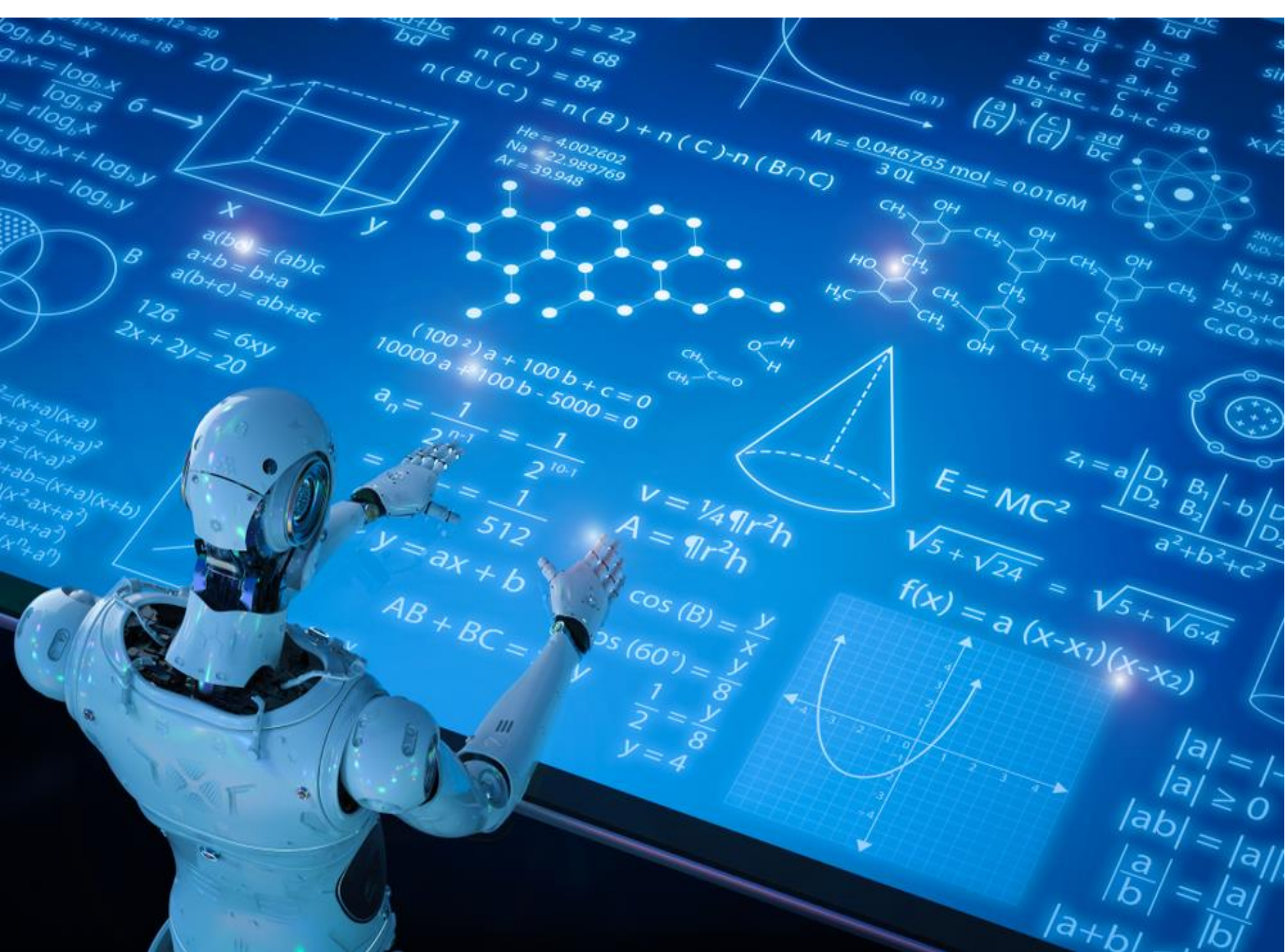
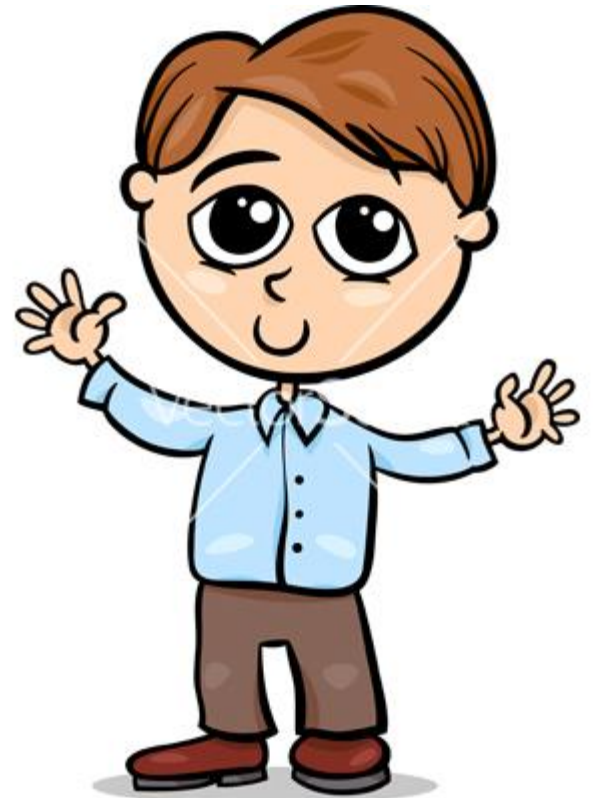
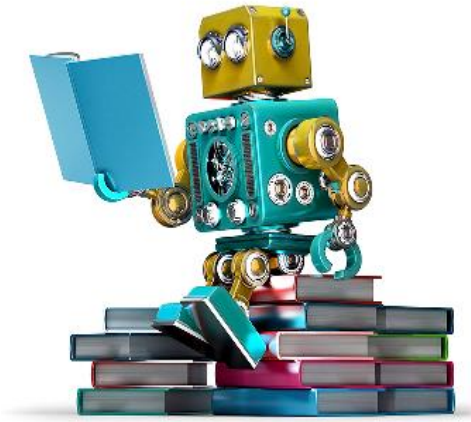


MACHINE LEARNING



www.chandanverma.com





Learn From Past Experiences

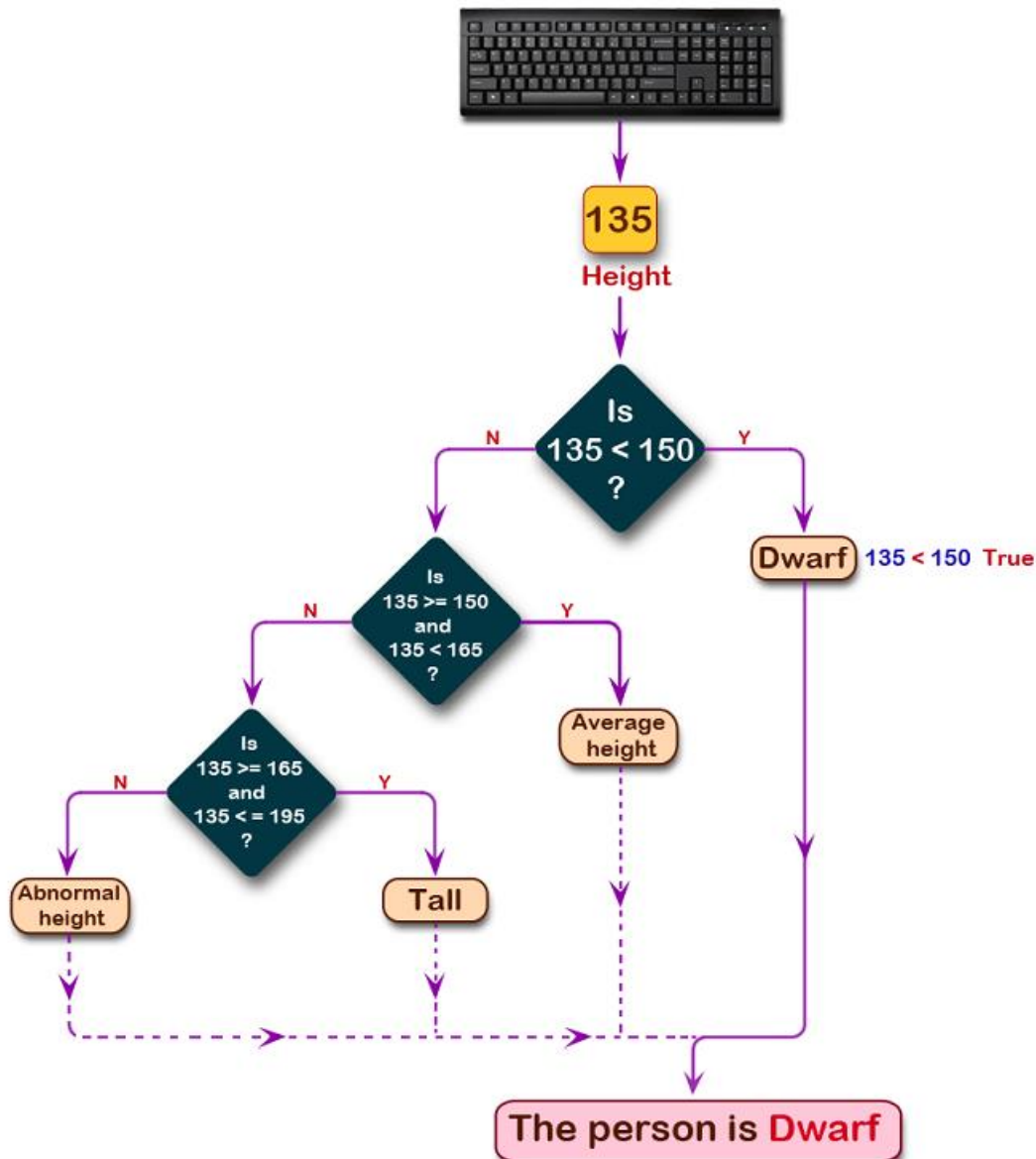
Follow the instruction



Given height of a person in centimeter and machine categorize them into

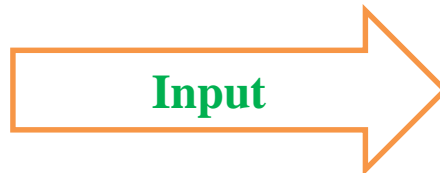
- 1. Tall**
- 2. Average**
- 3. Abnormal**
- 4. Dwarf**

Explicit Program



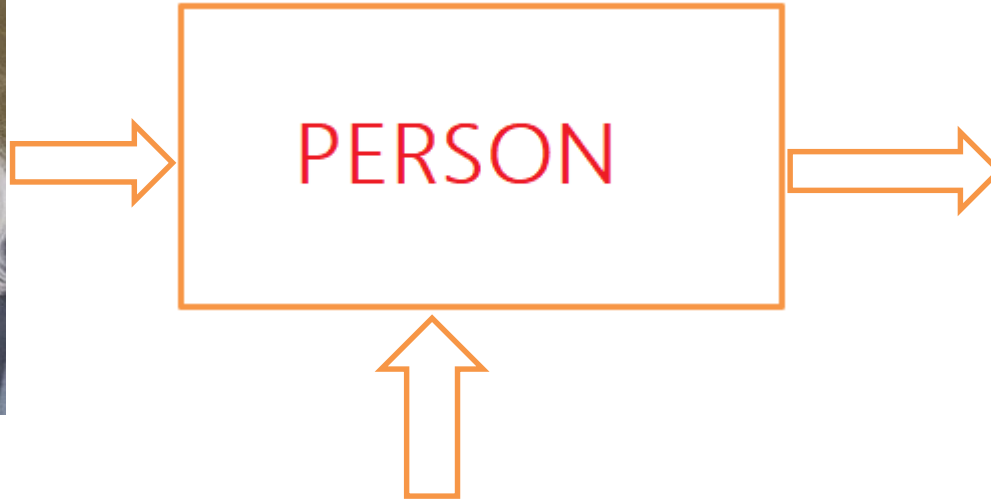
- Height < 150 → Dwarf
- Height = 150 → Average height
- Height >= 165 → Tall

RULE





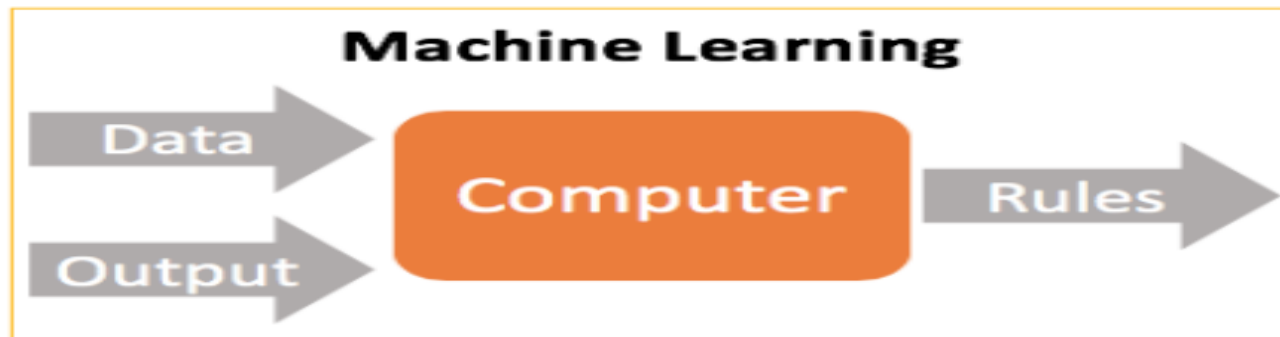
Julia Michaels



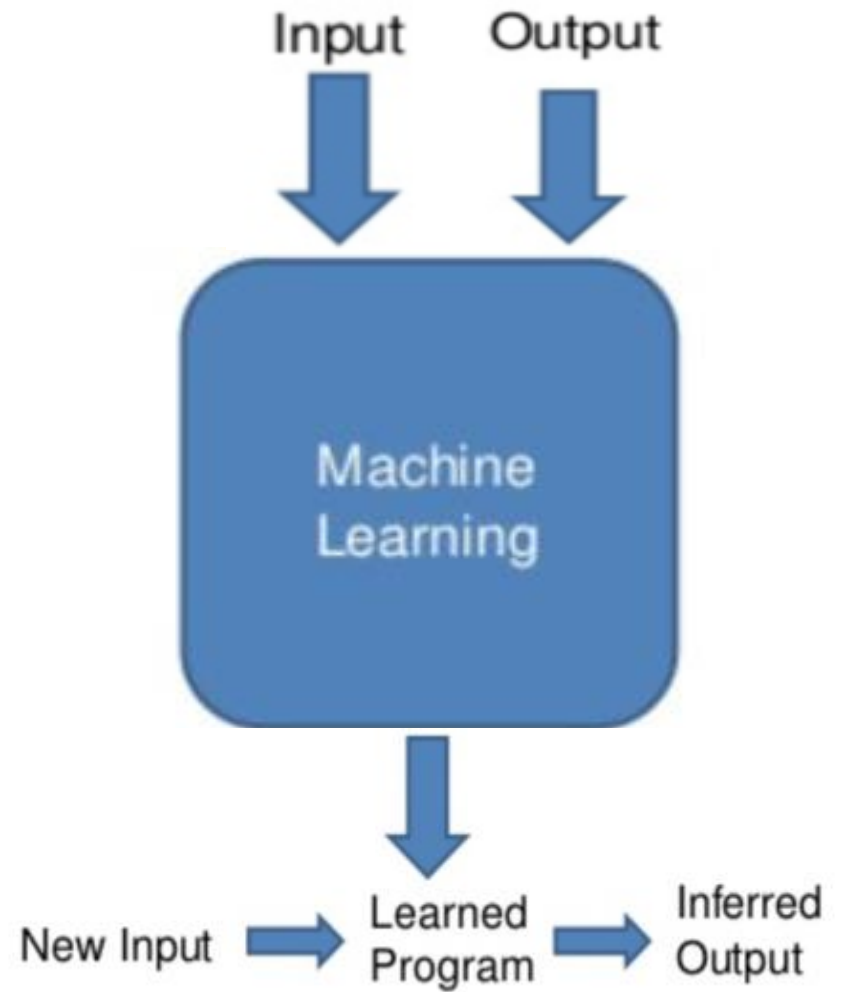
Look like
Anushka



Conventional Computer/Machine is based upon Certain rule and logic .



But today world is based on big data and need more complex and evidence based decision .



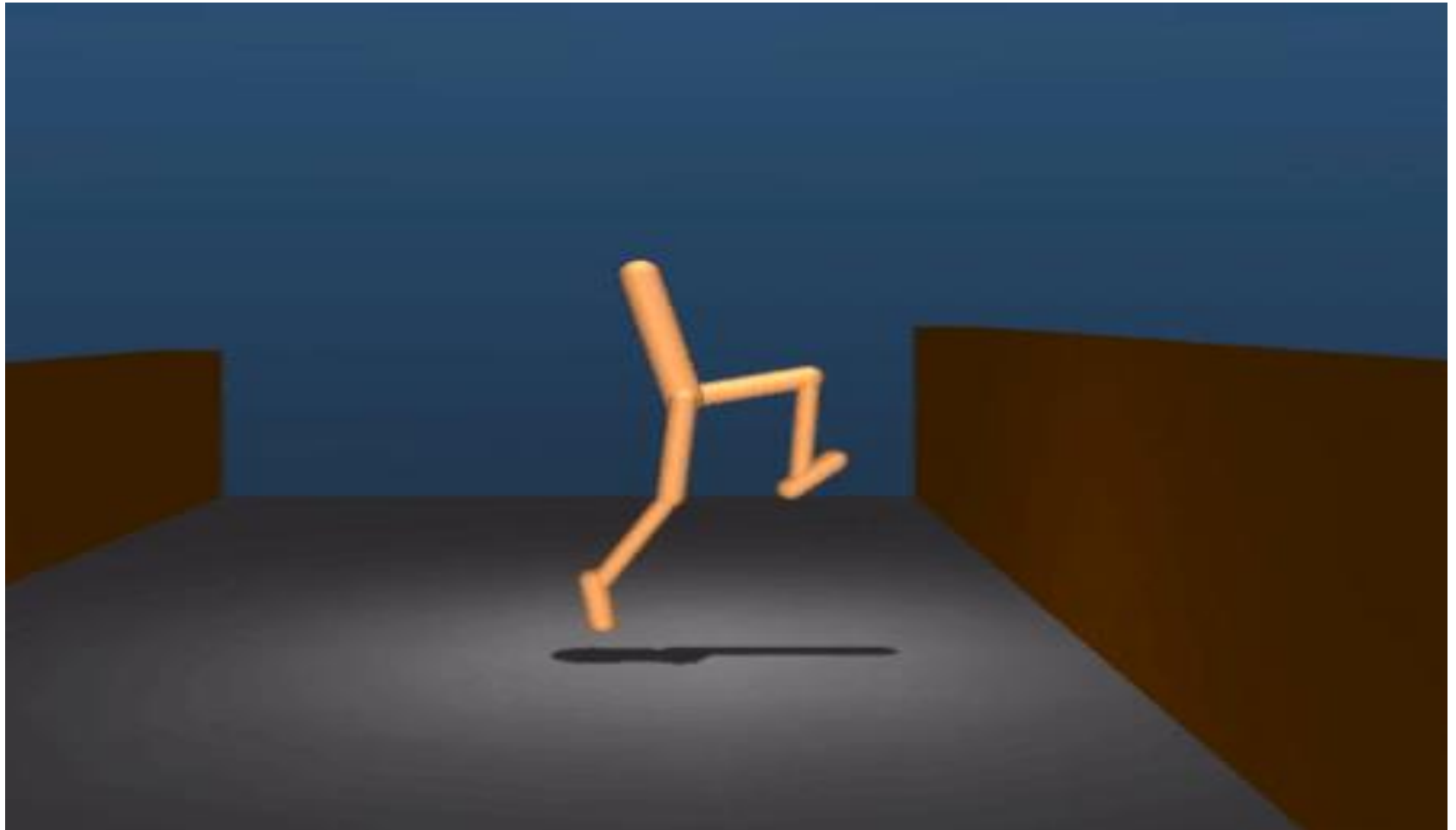
Machine Learning

Machine Learning is a subset of AI which enables the computer to act and make data-driven decisions to carry out a certain task. These programs or algorithms are designed in a way that they can learn and improve over time when exposed to new data.

Giving Computers a greater ability to understand information and to learn, to Reason , and act upon it

Making predictions based on experience

Learning From Past Experience



How Much Data is Generated Each Day

Here are some key daily statistics highlighted in the **infographic**:

- 500 million tweets are sent
- 294 billion emails are sent
- 4 petabytes of data are created on Facebook
- 4 terabytes of data are created from each connected car
- 65 billion messages are sent on WhatsApp
- 5 billion searches are made

By 2025, it's estimated that 463 exabytes of data will be created each day globally – that's the equivalent of 212,765,957 DVDs per day!

2019 *This Is What Happens In An Internet Minute*



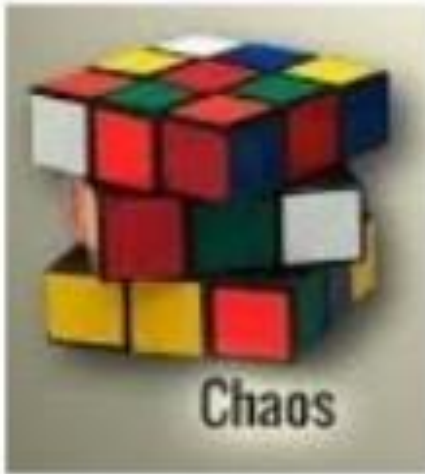
“Data Data everywhere, consumers are now more aware.

So mine the data with utmost care, and serve them everywhere.”

Data-Sciences



Data-Sciences



DATA MINING →



The Process of Discovering interesting and useful pattern and relationship is large volumes of data

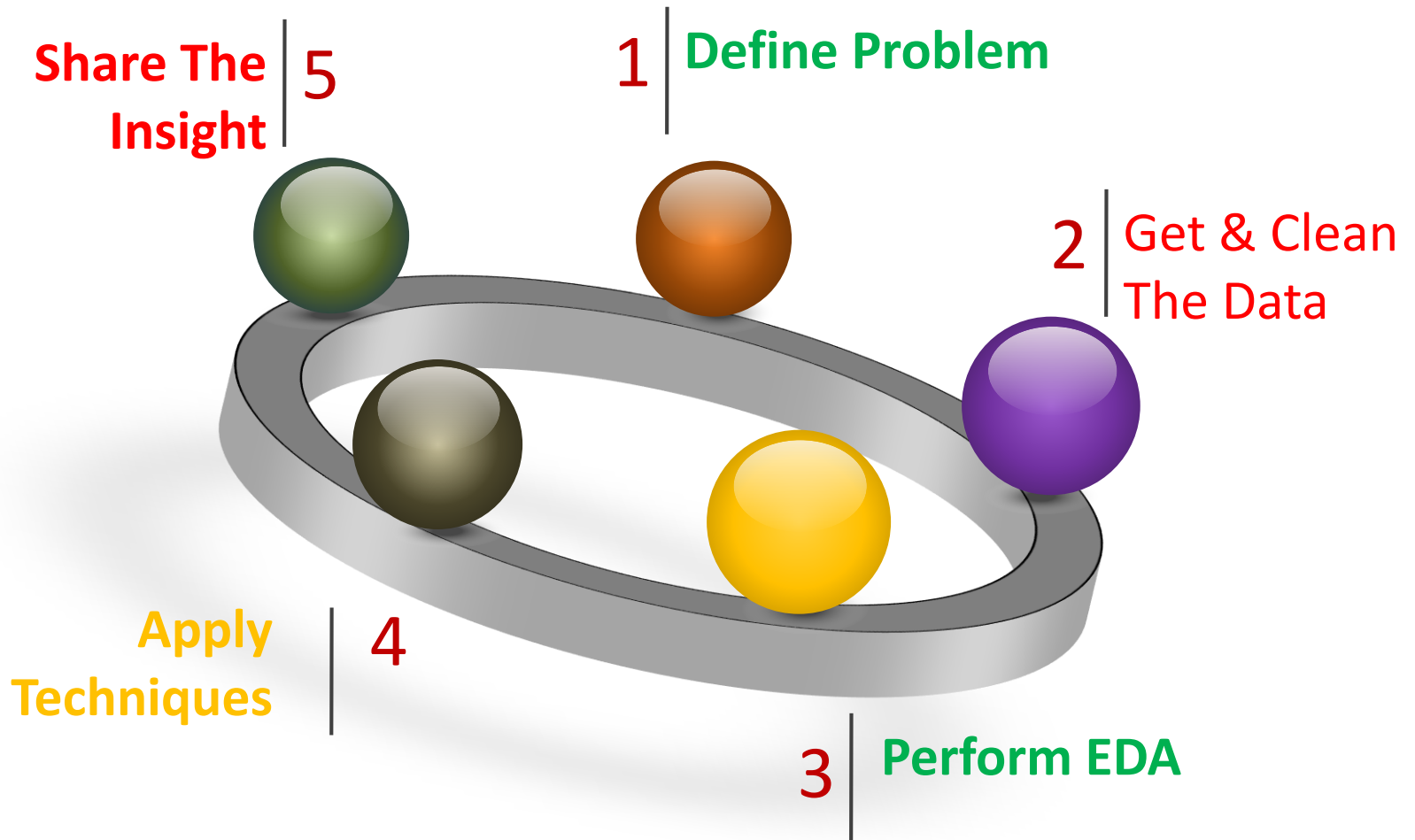
Data-Sciences

Using Data to Make Decision

Data science is the process of capturing data, processing it, communicating and analyzing it, and then maintaining it.

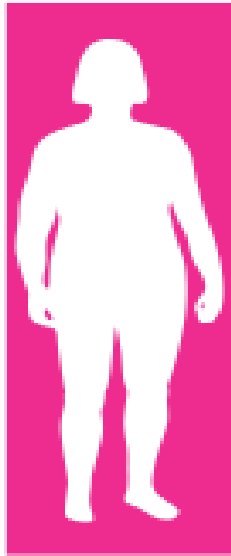
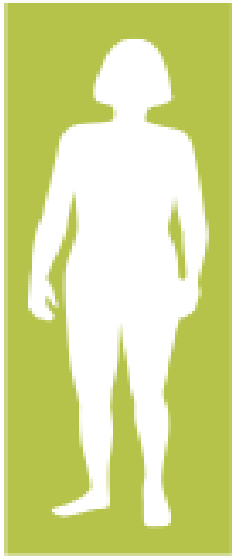
Data science is the process of analyzing data which involves applying Machine learning models, statistical models to derive insights and value from data.

Data-Science Workflow



Step-1 Define The Problem

My weight is w kg, Height , h cm ,
am I normal?



Step 2: Data – Collection

1	Gender	Height	Weight	Index
2	Male	174	96	4
3	Male	189	87	2
4	Female	185	110	4
5	Female	195	104	3
6	Male	149	61	3
7	Male	189	104	3
8	Male	147	92	5
9	Male	154	111	5
10	Male	174	90	3
11	Female	169	103	4
12	Male	195	81	2
13	Female	159	80	4
14	Female	192	101	3
15	Male	155	51	2
16	Male	191	79	2
17	Female	153	107	5
18	Female	157	110	5
19	Male	140	129	5
20	Male	144	145	5
21	Male	172	139	5

Step 3: Exploratory data analysis (EDA)



Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics, often with visual methods.

Univariate Analysis

- Box Plots
- Histogram

Multivariate analysis

- Scatter Plot
- Bar Chart
- Line plot
- Pie charts

When looking at a new dataset, **whether it is familiar to you or not**, it is important to use the following questions as guidelines for your preliminary analysis .

1. Is the data organized or not?
2. What does each row represent?
3. What does each column represent?
4. Are there any missing data points?
5. Do we need to perform any transformations on the columns?

1	Gender	Height	Weight	Index
2	Male	174	96	4
3	Male	189	87	2
4	Female	185	110	4
5	Female	195	104	3
6	Male	149	61	3
7	Male	189	104	3
8	Male	147	92	5
9	Male	154	111	5
10	Male	174	90	3
11	Female	169	103	4
12	Male	195	81	2
13	Female	159	80	4
14	Female	192	101	3
15	Male	155	51	2
16	Male	191	79	2
17	Female	153	107	5
18	Female	157	110	5
19	Male	140	129	5
20	Male	144	145	5
21	Male	172	139	5

Gender : Male / Female

Height : Number (cm)

Weight : Number (Kg)

Index :

0 - Extremely Weak

1 - Weak

2 - Normal

3 - Overweight

4 - Obesity

5 - Extreme Obesity

Step 4: Apply the Techniques

Here the Machine Learning Technique come into Picture to solve the problem.

Step 5: Share the Insights

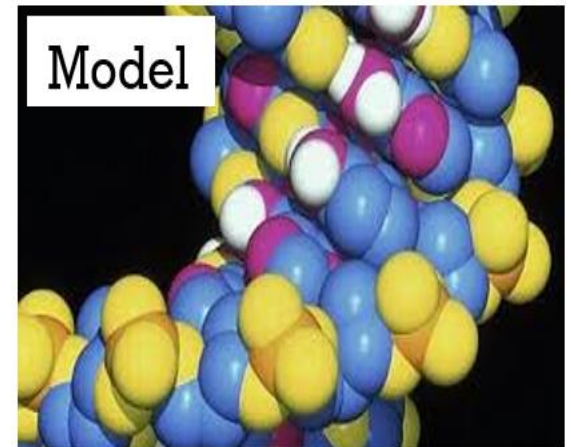
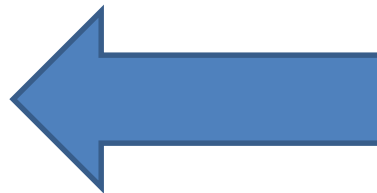
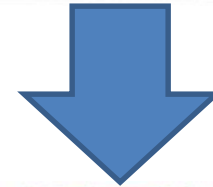
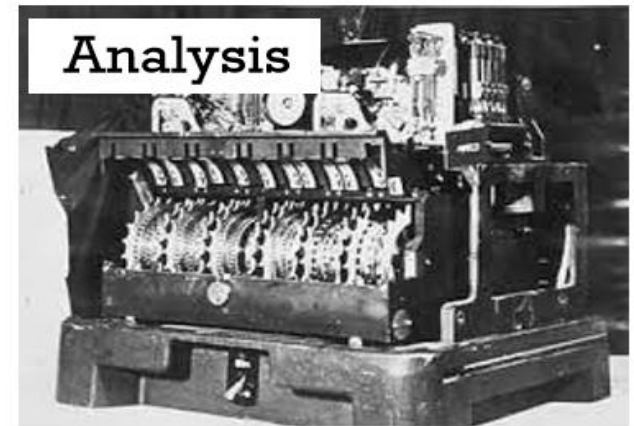
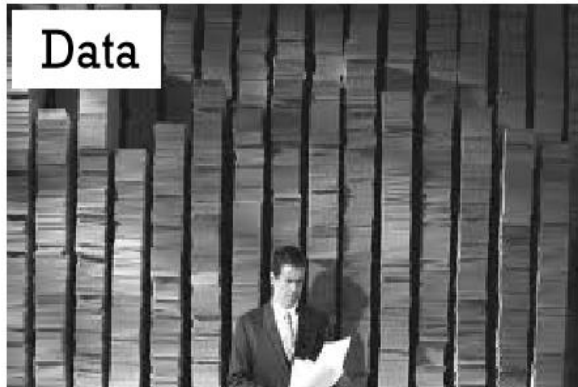
How to solve a problem in Data Science

Problems in Data Science are solved using Algorithms. But, the biggest thing to judge is which algorithm to use and when to use it?

Machine Learning Algorithms



solving a problem in Data Science



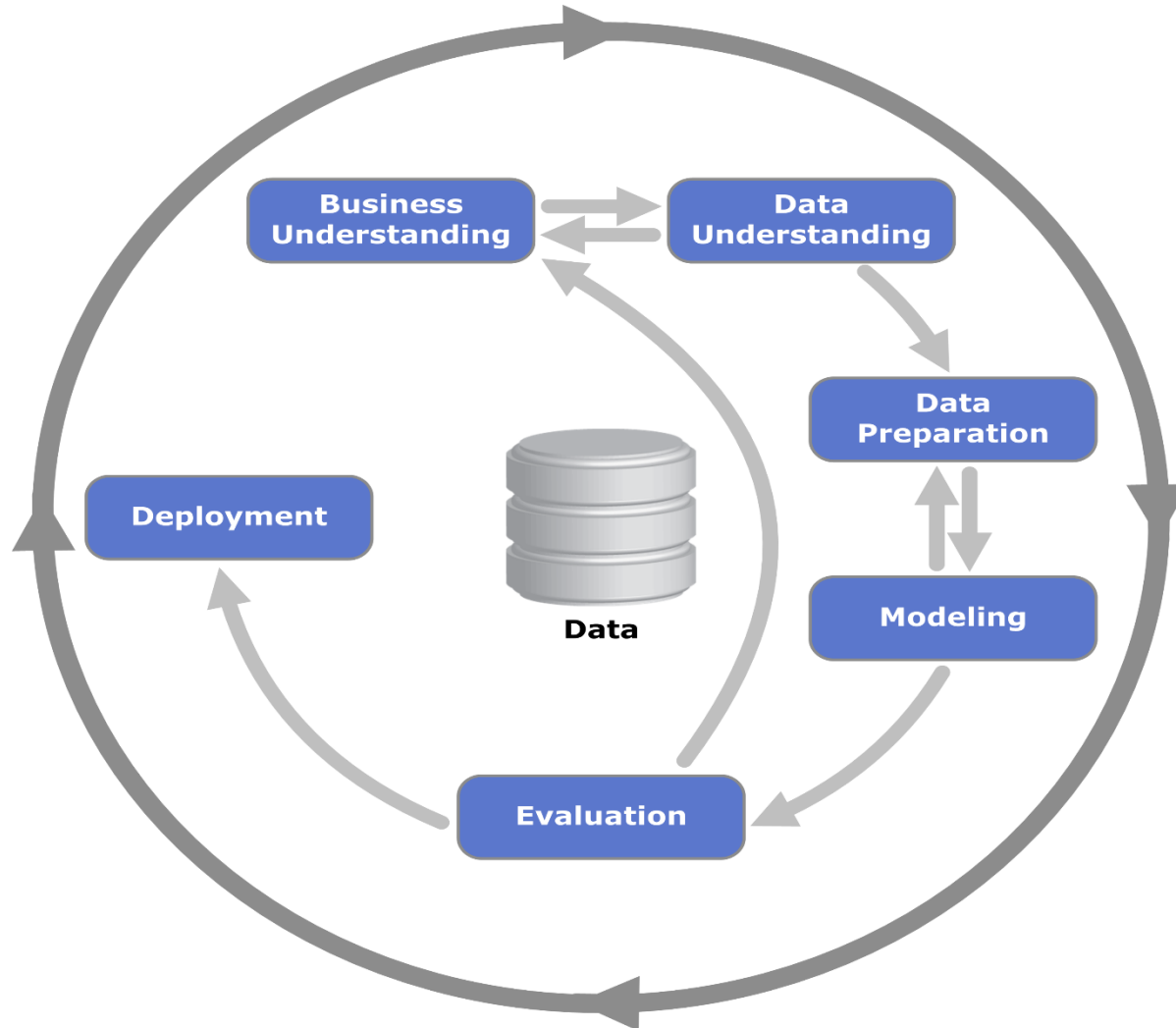
How to solve Machine Learning problems?

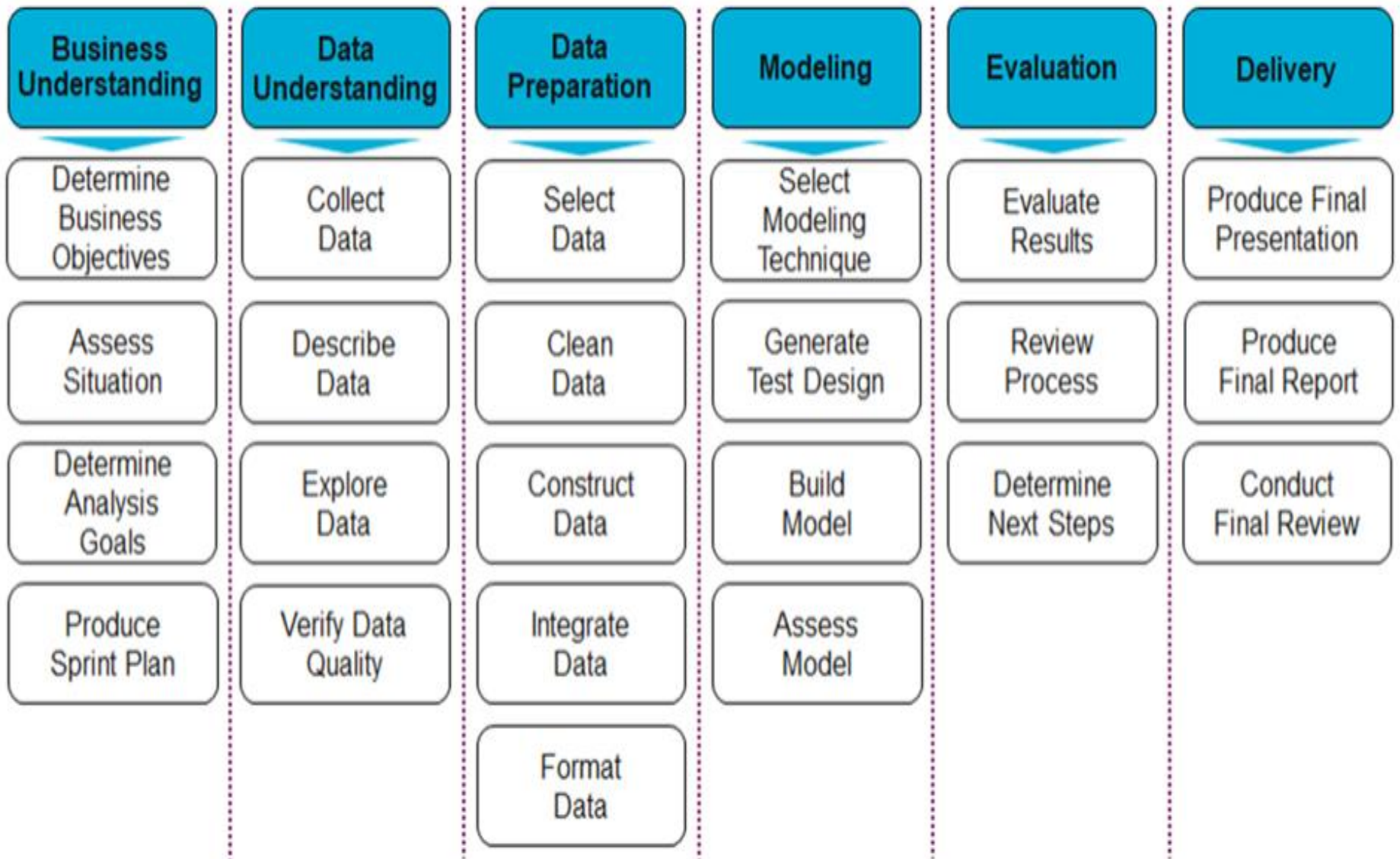
CRISP-DM: Data mining methodology to investigate (Big) Data
Cross-Industry Standard Process for Data Mining

1. **Business Understanding**
2. **Data Understanding**
3. **Data preparation**
4. **Modeling**
5. **Evaluation**
6. **Deployment**
7. **Start again in Iterative process**

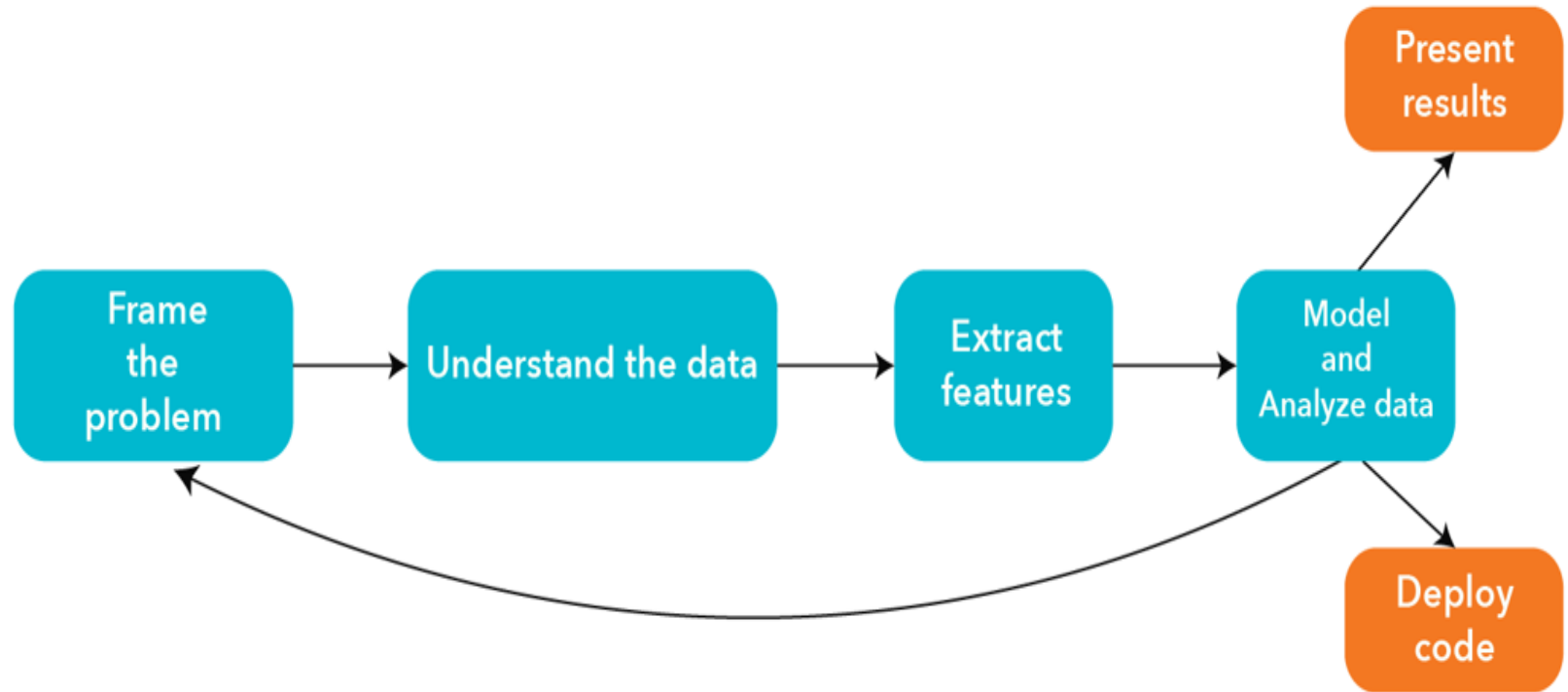
CRISP-DM

Cross-Industry Standard Process for Data Mining





Data Science Road Map



Machine Learning

Problems in Data Science are solved using Algorithms.

Machine Learning is Problem Solving Algorithms

But, the biggest thing to judge is

- **Which Type of Problem ?**
- **What type of Algorithm for Given Problem?**

Type of Problem

1. Is this fruit Sweet or Sour ?
2. Is this Weird?
3. How much or many ?
4. how is this organized?
5. What should i do next ?

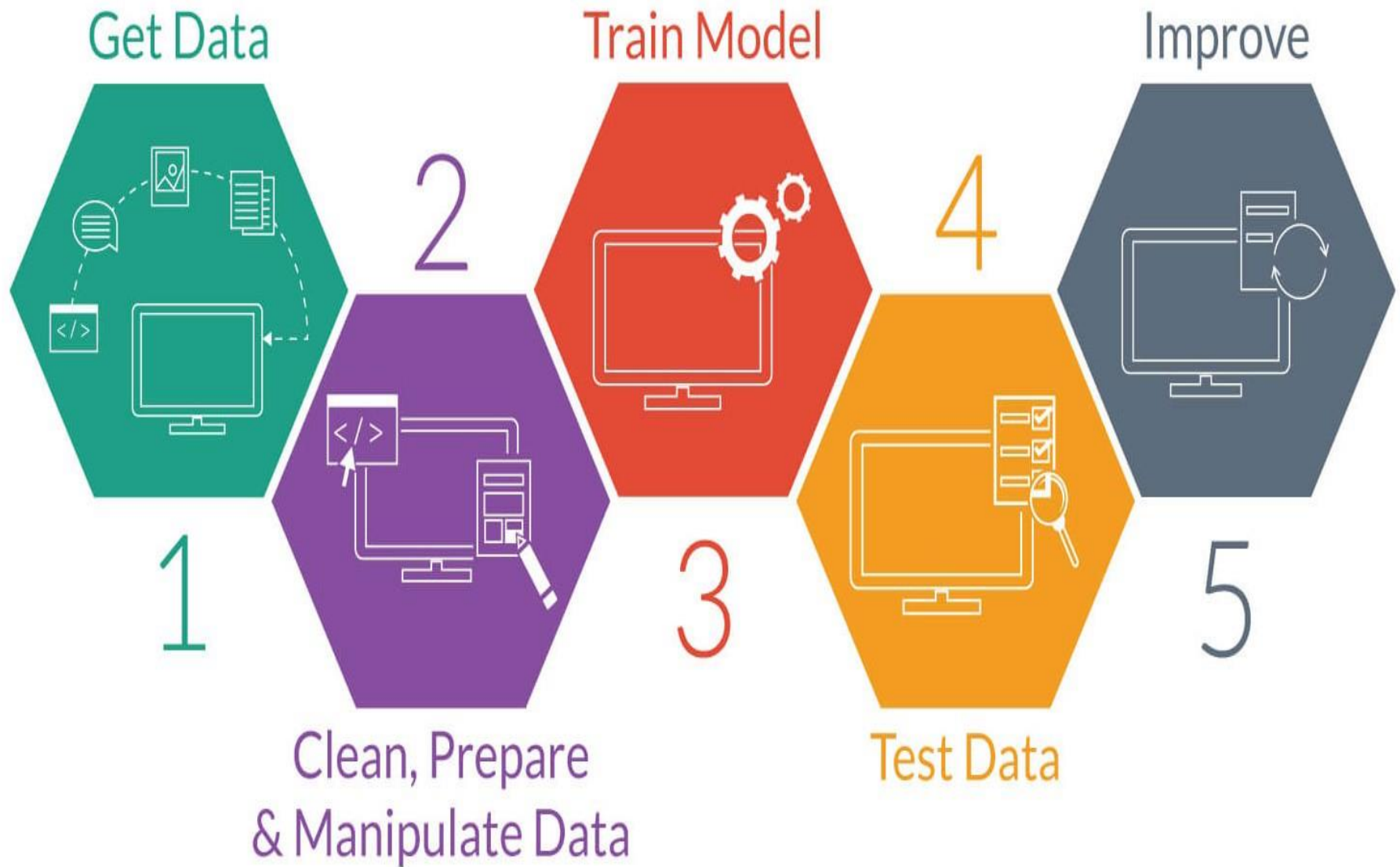


Machine Learning = Representation + Evaluation + Optimization

Machine learning constructs algorithms that can learn from data

Steps involved in machine learning

- **Data Collection**
- **Preprocessing**
- **Data Split**
- **Training the system**
- **Testing the model**
- **Model storage and reuse**



Feature and Target

					features	target
samples (train)	type (category)	# rooms (int)	surface (float m2)	public trans (boolean)	sold (float k€)	
	Apartment	3	50	TRUE	450	
	House	5	254	FALSE	430	
	Duplex	4	68	TRUE	712	
	Apartment	2	32	TRUE	234	

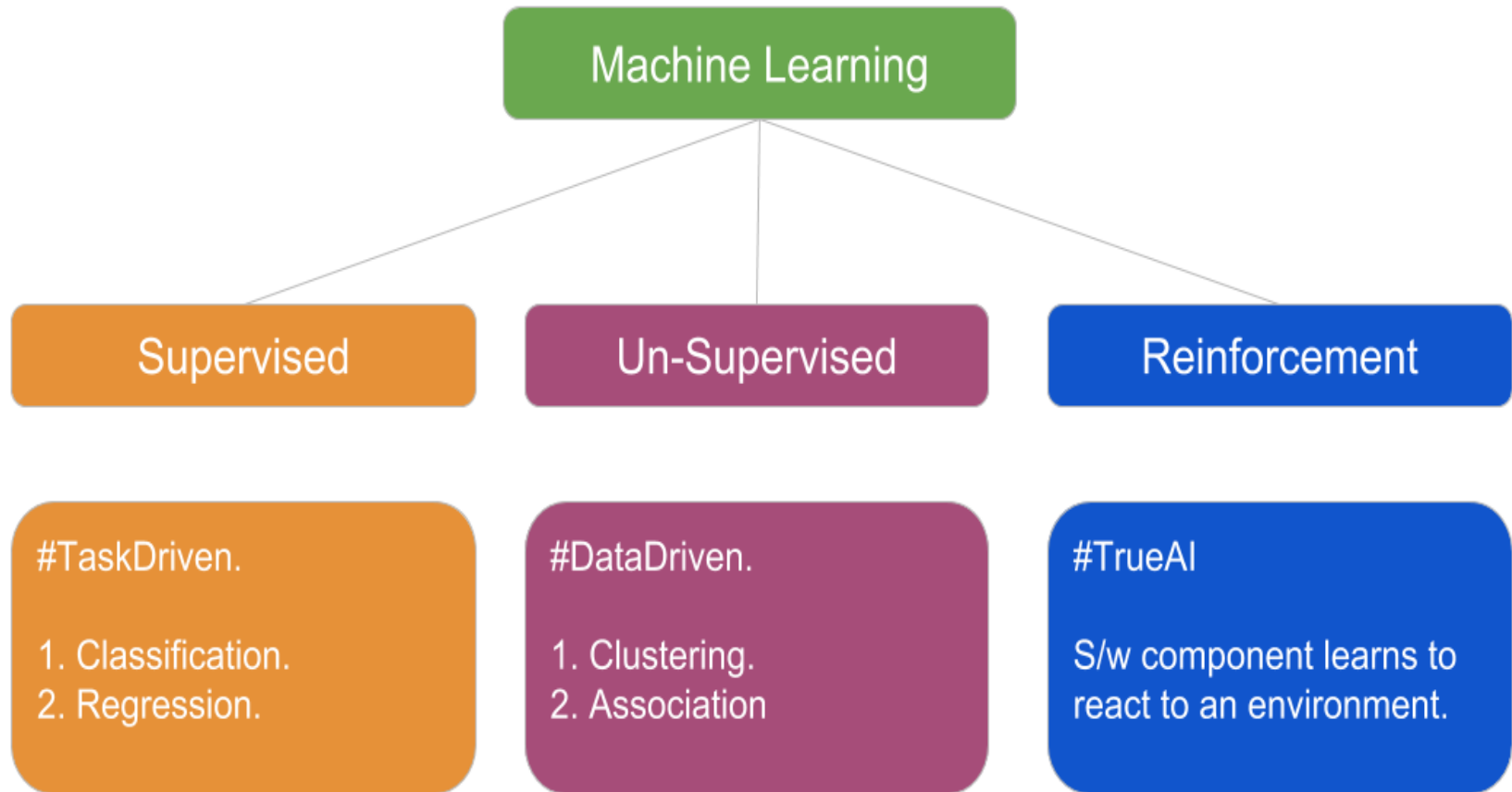
Training -Model

A **model** defines the relationship between features and label.

Training means creating or learning the model.
Training is the process of making the system able to learn.

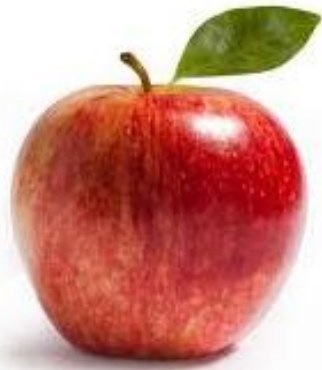
Inference means applying the trained model to unlabeled examples

Machine Learning-Algorithms



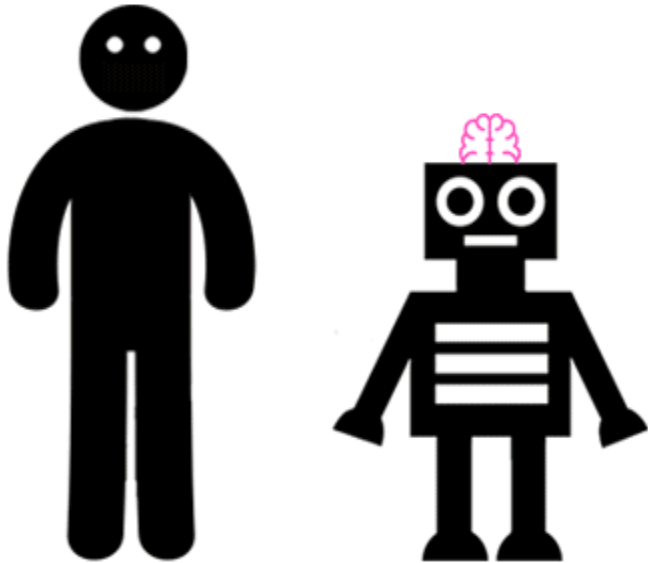
Supervised learning





Supervised learning is known as Predictive learning

Supervised learning



Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(X)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

Classification: predict a *discrete label*. The outputs fall under a finite set of possible outcomes.

Regression: predict outputs that are *continuous*. The outputs are quantities that can be flexibly determined based on the inputs of the model rather than being confined to a set of possible labels.

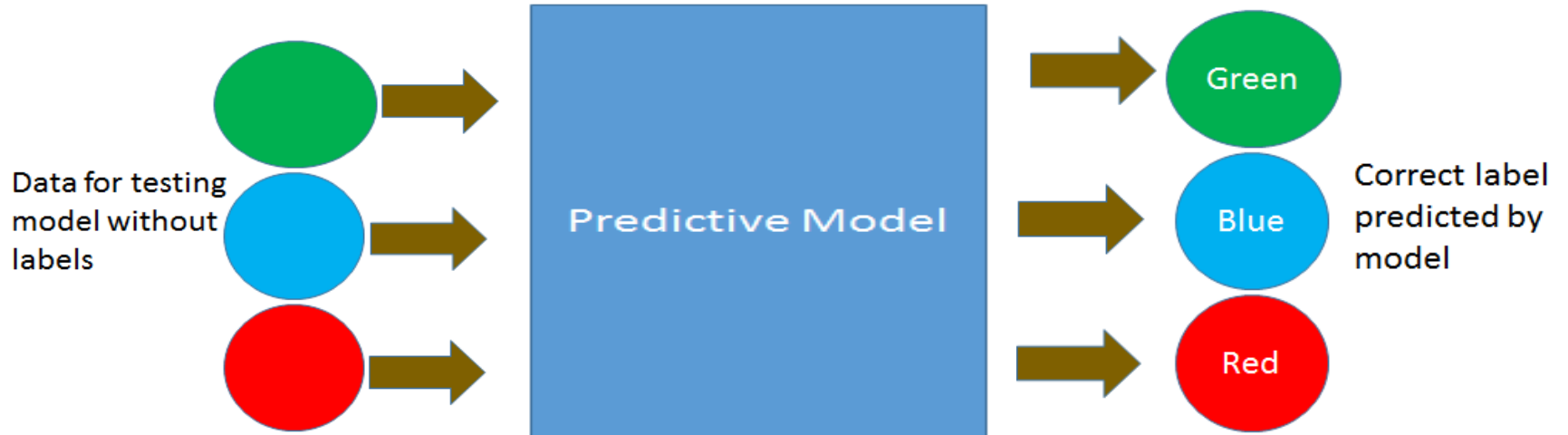
1. You already learn from your previous work about the physical characters of fruits.
2. Your previous work is called as **training data** in data mining.
3. so you already learn the things from your train data, this is because of **response variable**.
4. Response variable mean just a **decision variable**.

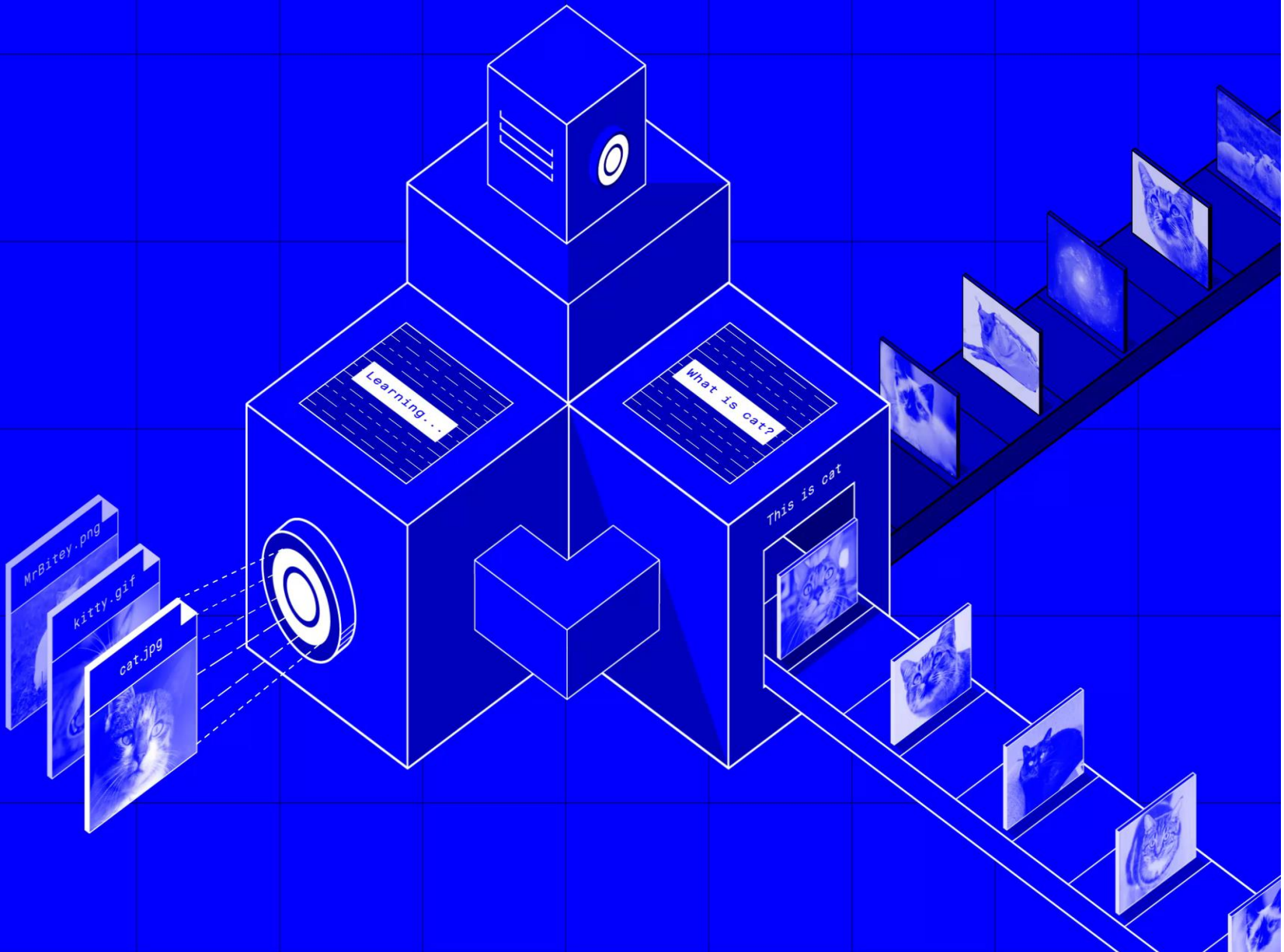
NO.	SIZE	COLOR	SHAPE	FRUIT NAME
1	Big	Red	Rounded shape with a depression at the top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curving cylinder	Banana
4	Small	Green	Round to oval,Bunch shape Cylindrical	Grape

Training Phase



Testing Phase





REGRESSION



?

CLASSIFICATION



?

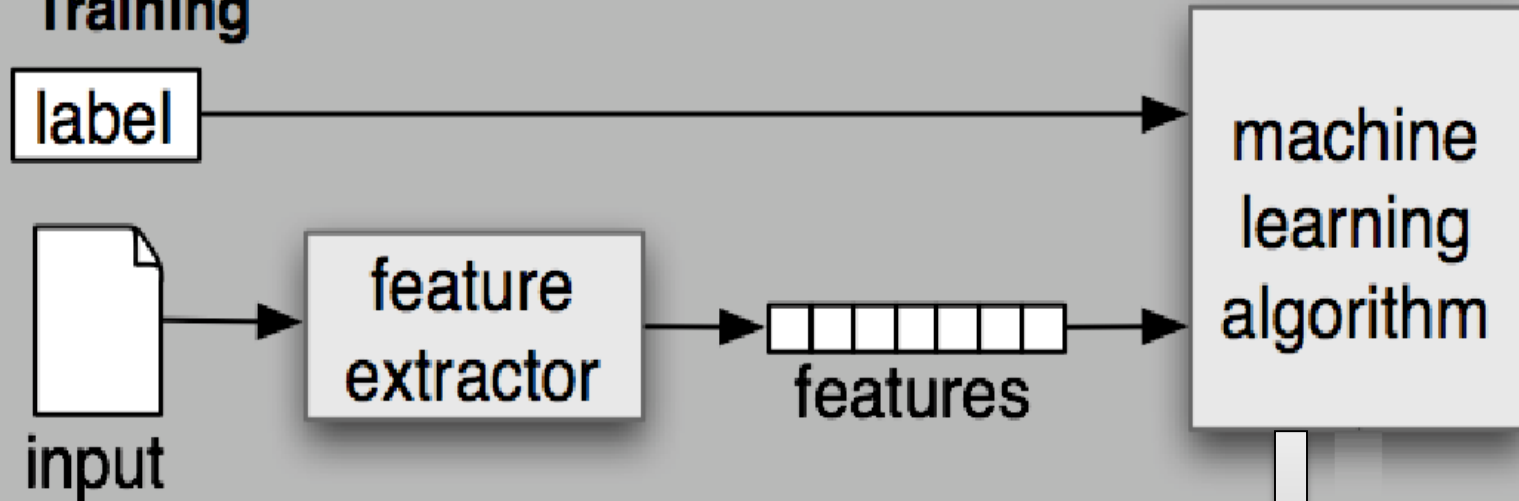
S

M

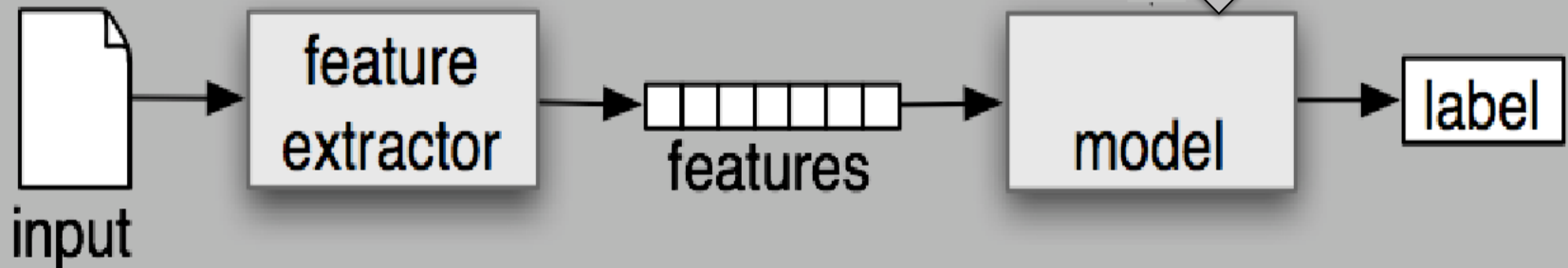
L

ML PIPELINE (SUPERVISED)

Training



Prediction



Regression or classification

1. Is this cancer?
2. Will this person like this movie?
3. What is market Value of this house?
4. What is the value of a house in California?
5. What is the probability that a user will click on this ad?
6. Is a given email message spam or not spam?
7. Is this an image of a dog, a cat, or a hamster?

Regression or classification

- Predict whether an email is spam or not
- Predict whether it will rain or not
- Predict whether a user is a power user or a casual user
- Predict the height of a potted plant from the amount of rainfall
- Predict salary based on someone's age and availability of high-speed internet
- Predict a car's MPG (miles per gallon) based on size and model year

Machine Learning (ML) applications

- **Prediction**
- **Classification**
- **Recognition**

Recommendations

Process of learning preferences of people over time and recommend new things which might not have been known to that person.

Amazon recommends products to you, this is possible because of machine learning

Unsupervised learning / Descriptive models:

There are **no labels associated** with data points. These machine learning algorithms organize the data into a group of clusters to describe its structure and make complex data look simple and organized for analysis.

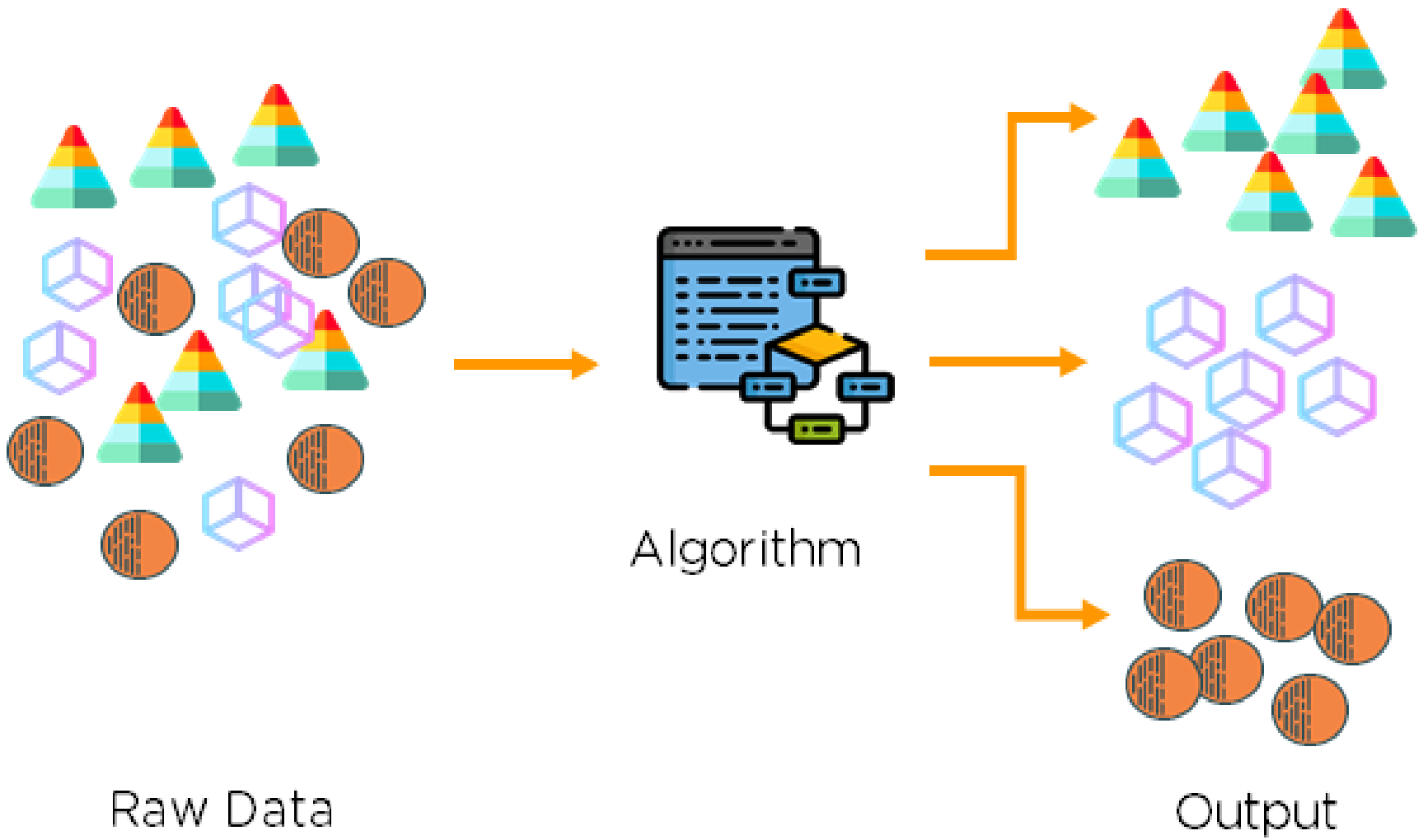
It is used to **draw inferences from datasets** consisting of input data without labeled responses.

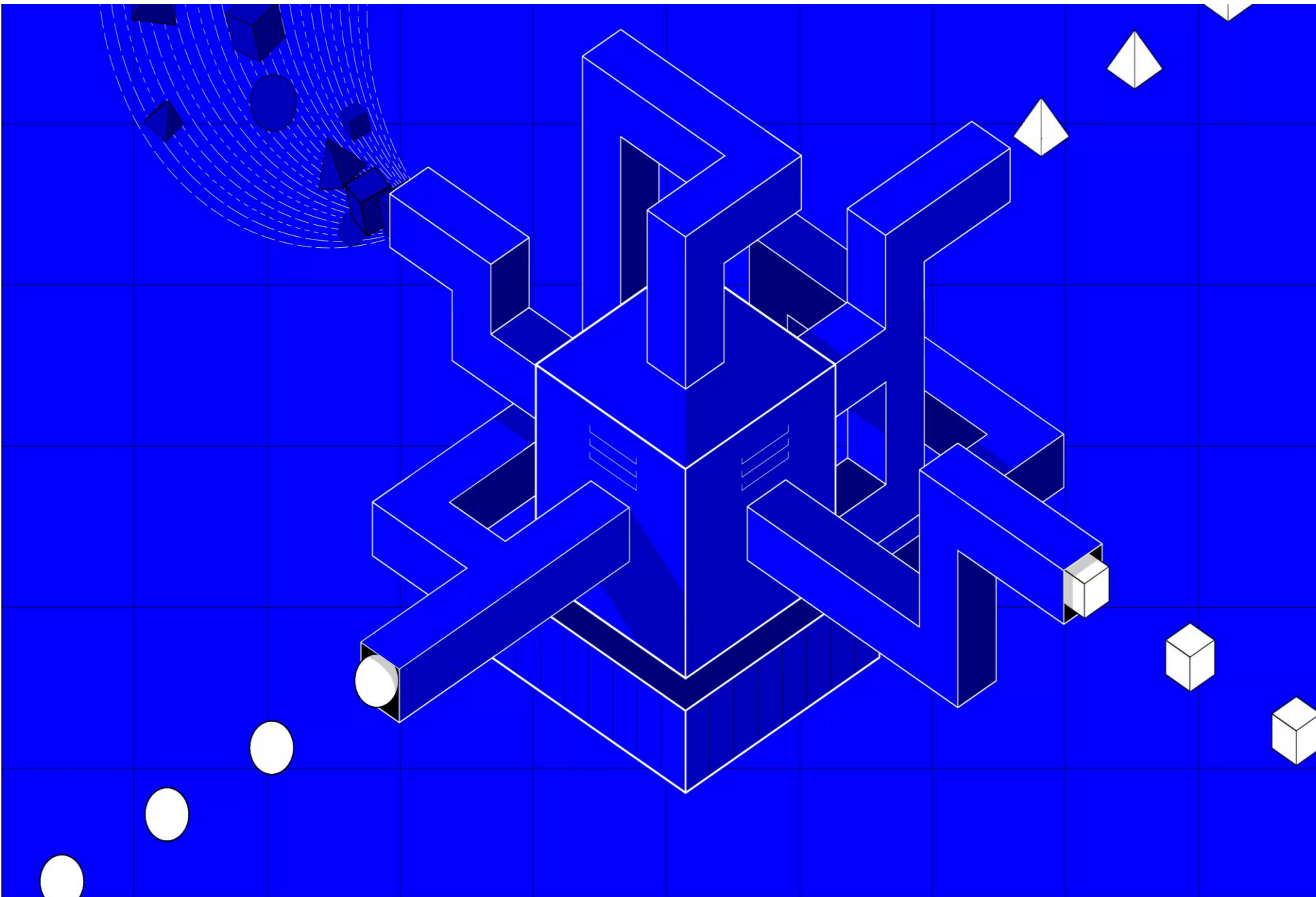
the machine learns through observation & find structures in data

Unsupervised learning / Descriptive models:

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

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





INPUT

"OK, let me try..."

#2

#1

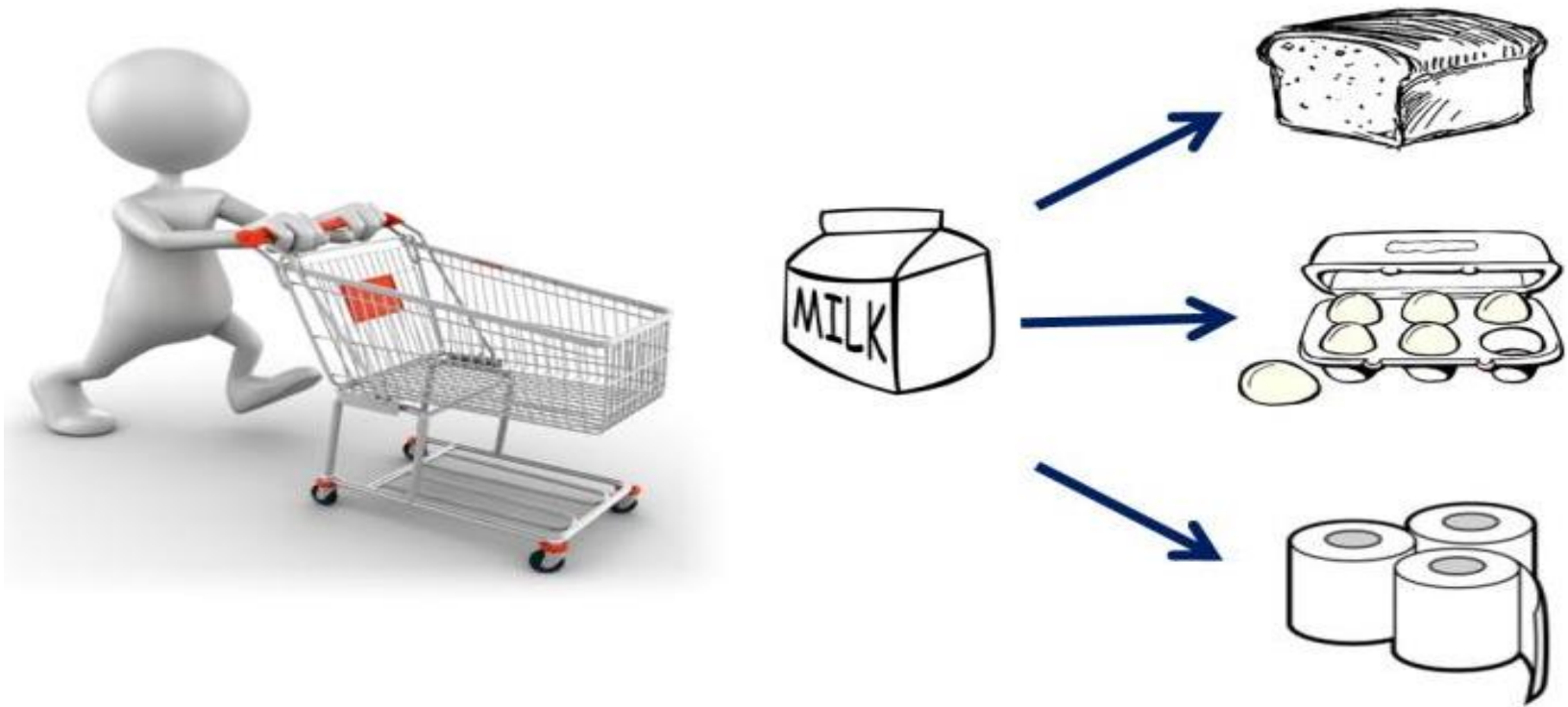
Medium size, 4 wheels, glass...

Small size, 2 wheels...

Big size, 4+ wheels, glass...

UNSUPERVISED LEARNING

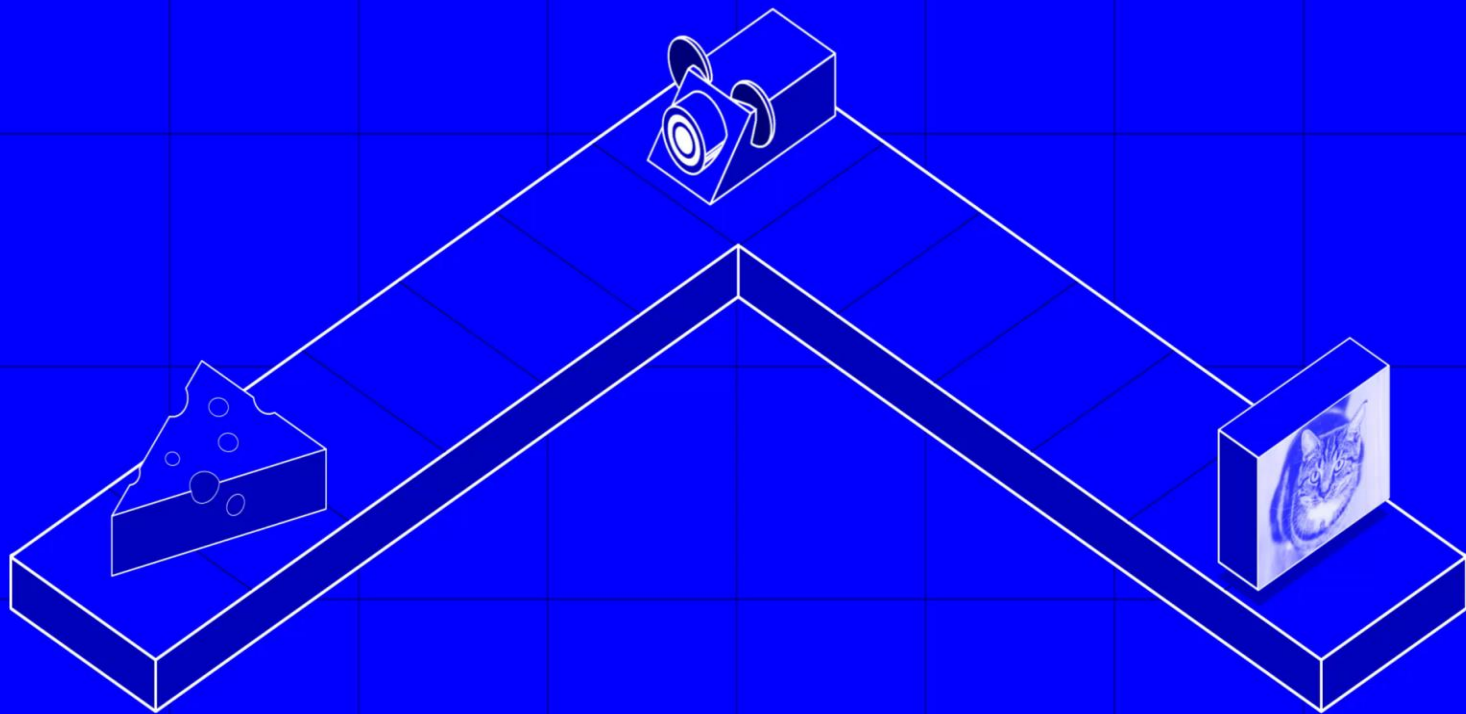
Association

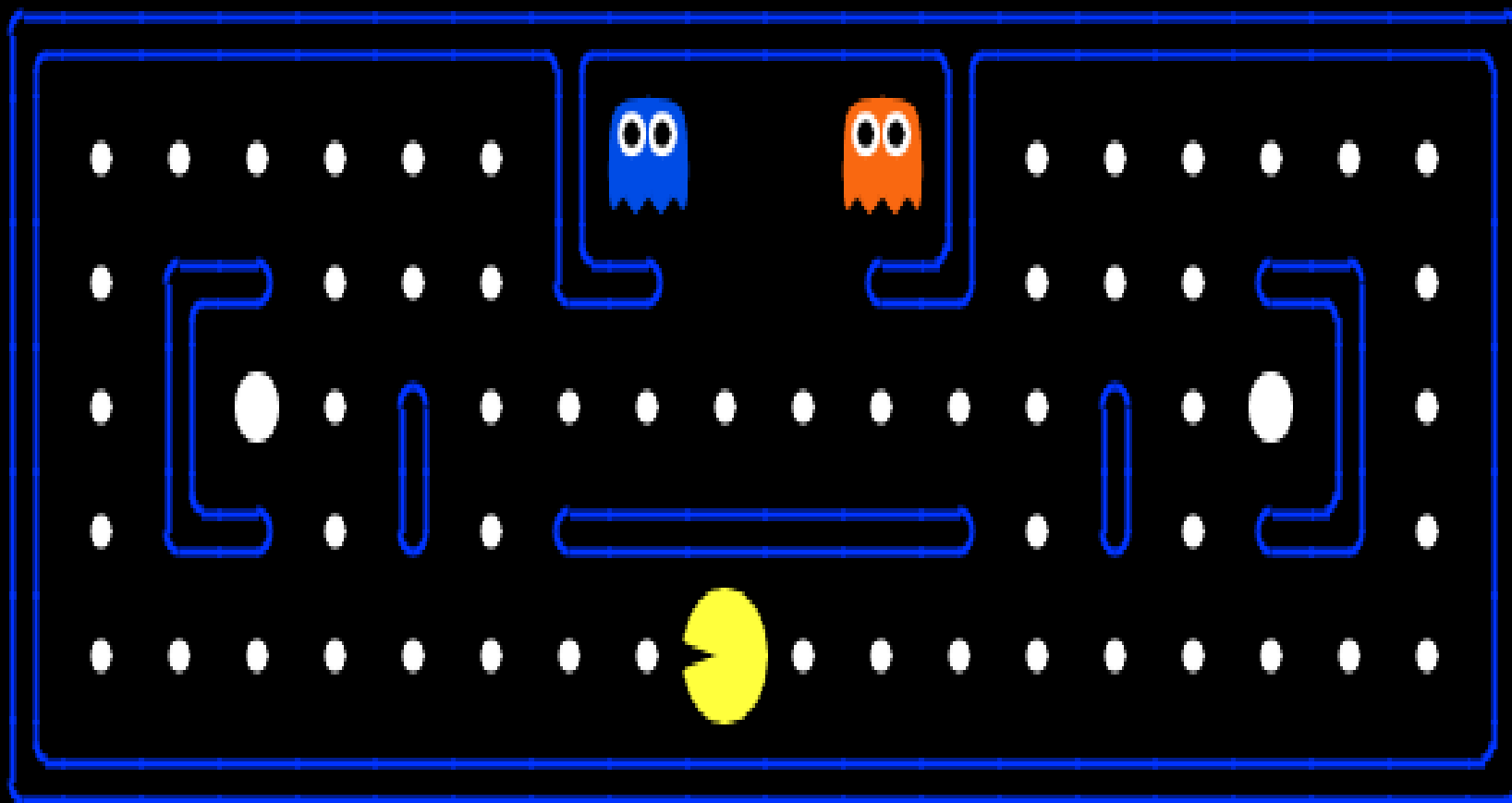


Reinforcement Learning

It is the ability of an agent to interact with the environment and find out what is the best outcome. It follows the concept of hit and trial method. The agent is rewarded or penalized with a point for a correct or a wrong answer, and on the basis of the positive reward points gained the model trains itself. And again once trained it gets ready to predict the new data presented to it







SCORE: 0

- Labeled data
- Direct feedback
- Predict outcome/future



- No labels
- No feedback
- “Find hidden structure”

- Decision process
- Reward system
- Learn series of actions



MACHINE LEARNING



SUPERVISED LEARNING



UNSUPERVISED LEARNING



CLASSIFICATION



REGRESSION



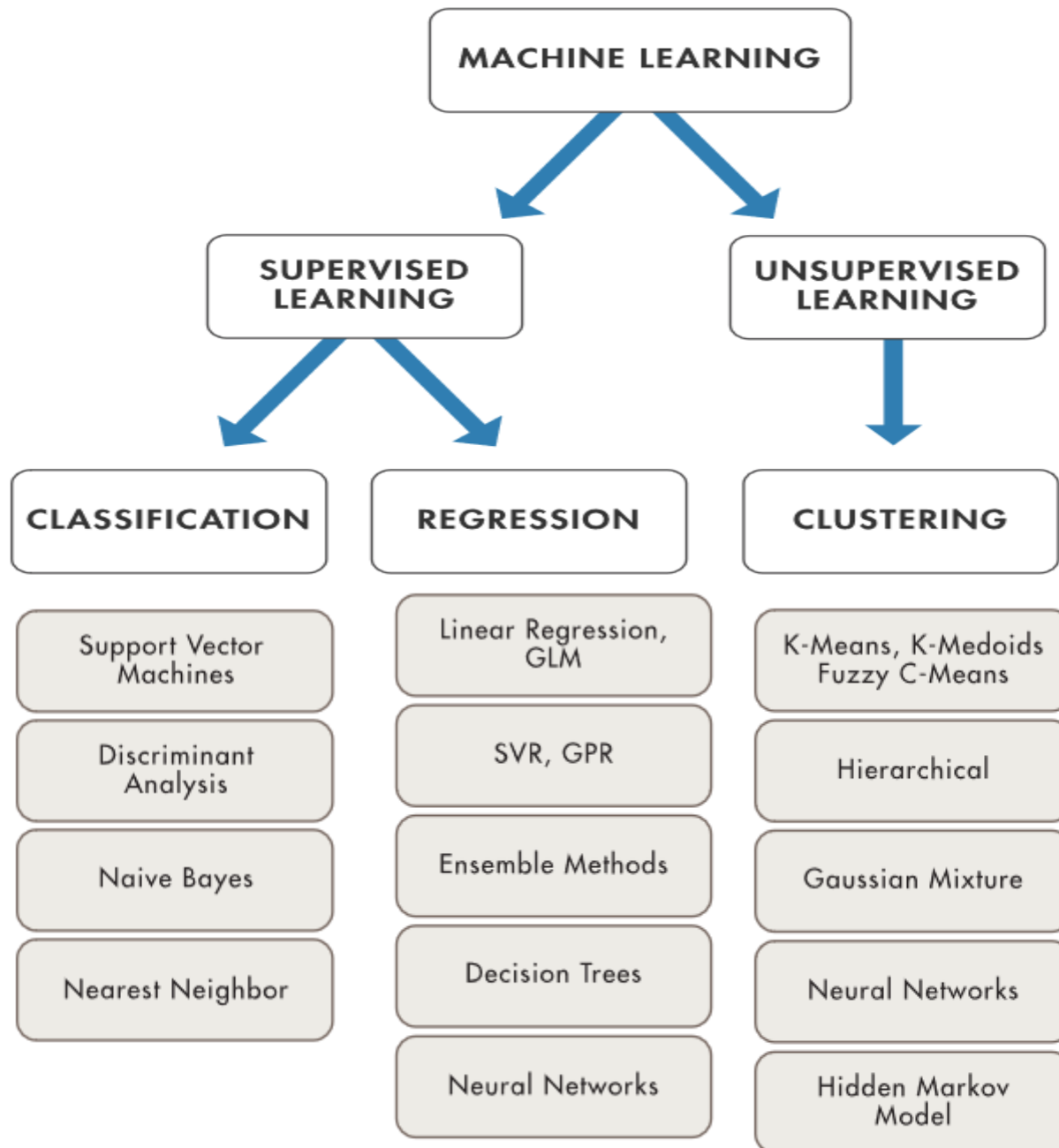
CLUSTERING



ASSOCIATION

List of Common Machine Learning Algorithms

- **Naïve Bayes Classifier Algorithm**
- **K Means Clustering Algorithm**
- **Support Vector Machine Algorithm**
- **Apriori Algorithm**
- **Linear Regression**
- **Logistic Regression**
- **Artificial Neural Networks**
- **Random Forests**
- **Decision Trees**
- **Nearest Neighbours**
- **self-organizing maps,**





Machine learning



Supervised



Unsupervised



Reinforcement



Regression

- Linear
- Polynomial



Decision Tree



Random forest



Classification

- KNN
- Trees
- Logistic Regression
- Naive-Bayes
- SVM



Clustering

- SVD
- PCA
- K-means



Association analysis

- Apriori
- FP-Growth

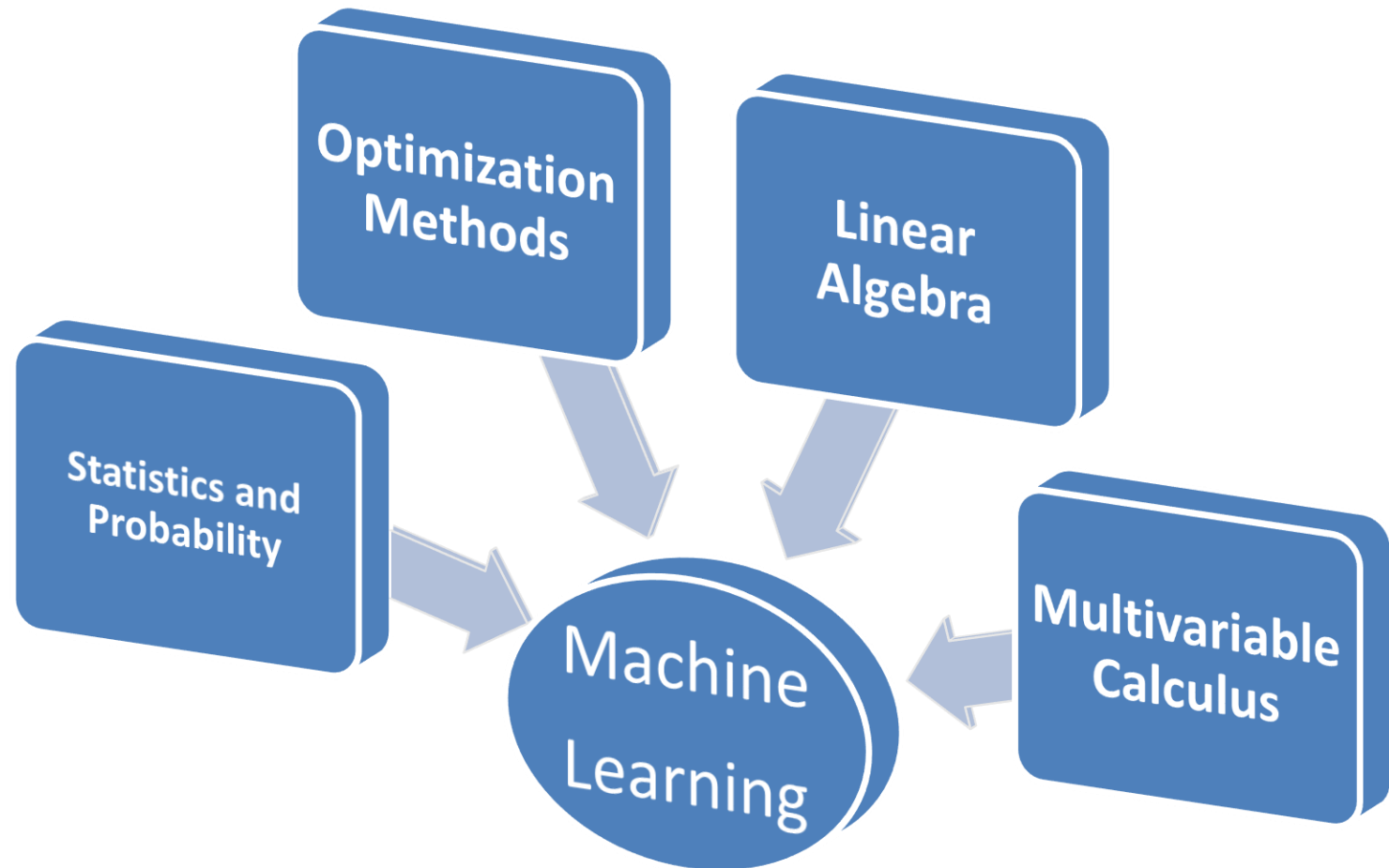


Hidden Markov Model

Continuous

Categorical

Essential Math Skills for Machine Learning



Statistics and Probability

- Mean
- Median
- Mode
- Standard deviation
- variance
- Correlation
- covariance
- distributions
- p-value
- Baye's Theorem
- A/B Testing
- Monte Carlo Simulation
- Precision
- Recall
- Confusion Matrix

Multivariable Calculus

- Functions of several variables
- Derivatives and gradients
- Step function
- Sigmoid function
- Logit function
- ReLU function
- Cost function
- Plotting of functions
- Minimum
- Maximum

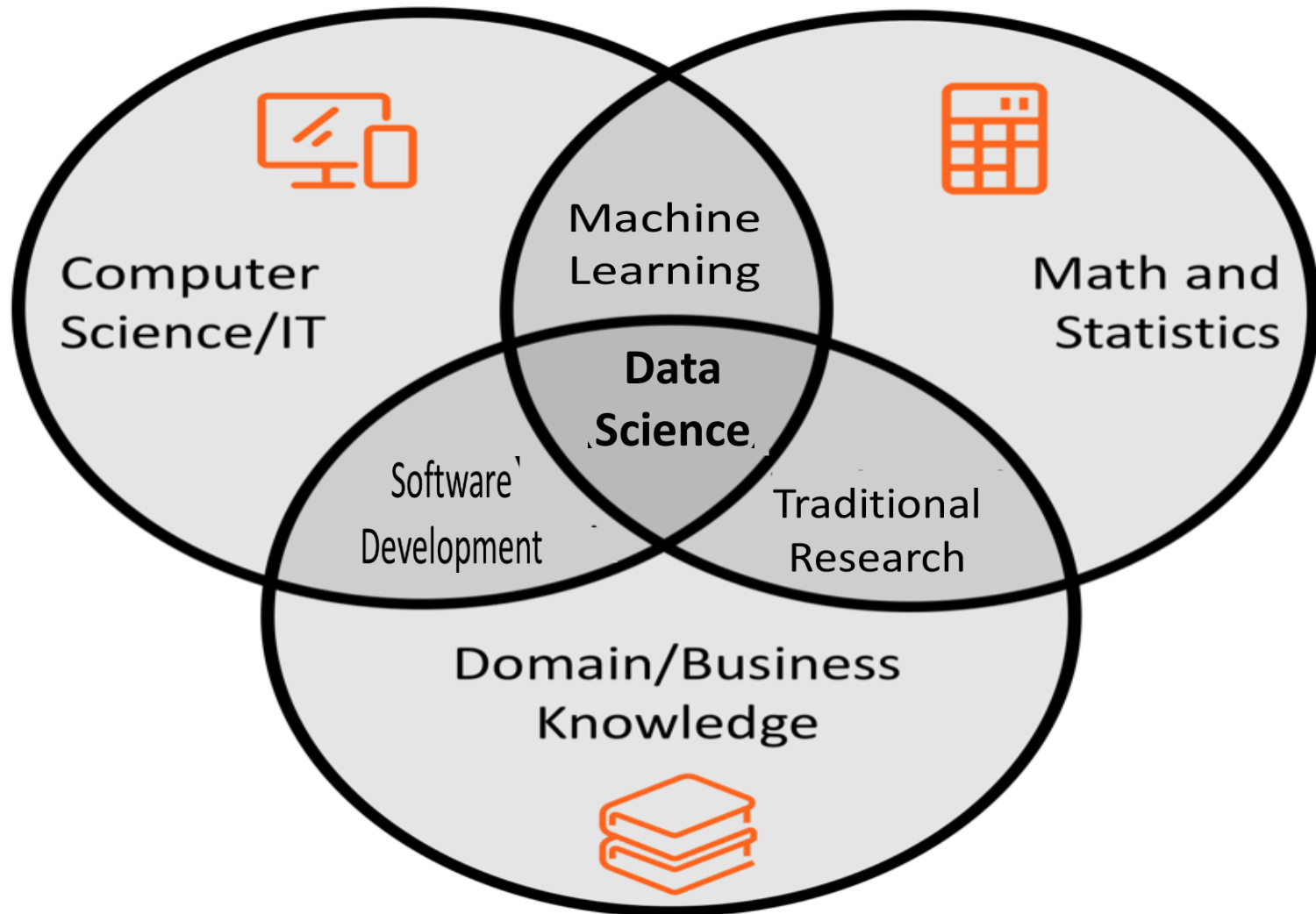
Linear Algebra

- Vectors
- Matrices
- Transpose of a matrix
- The inverse of a matrix
- The determinant of a matrix
- Dot product
- Eigenvalues
- Eigenvectors

Optimization Methods

- Cost function
- Objective function
- Likelihood function
- Error function
- Gradient Descent
- Stochastic GDO

competencies that Machine Learning requires



Most Popular Machine Learning Software Tool



 PyTorch

 Keras



Tensorflow

colab



Spark
MLlib

 Microsoft
CNTK



Data

- Structured data
- Unstructured data
- Qualitative data
- Quantitative data
- Discrete data
- Continuous data
- Nominal
- Ordinal
- Interval

Tool

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Sklearn
- R
- Tensorflow
- keras
- NLTK
- Request
- Beautiful Soup
- Pickle

Math

- Probability Theory and Statistics
- Bayes' Theorem
- Random Variables
- Variance and Expectation
- Conditional and Joint Distributions
- Standard Distribution
- Calculus

Algorithms

- Supervised
- Unsupervised
- Semi-Supervised
- Reinforcement

Roles-In Machine Learning Domain

Data Scientist - builds your Machine Learning Models.

Data Engineer - deploys and monitors your Machine Learning Models. Also works with data integration.

Data Visualization - creates dashboards and data insight.

Process Owner - manages the team, manages stakeholder expectations and maintains a vision.

- Prove / disprove hypotheses.
- Information and Data gathering.
- Data wrangling.
- Algorithm and ML models.
- Communication.

Data Scientist



- Build Data Driven Platforms.
- Operationalize Algorithms and Machine Learning models.
- Data Integration.
- Monitoring.

Data Engineer



- Storytelling.
- Build Dashboards and other Data visualizations.
- Provide insight through visual means.

Visualization Expert

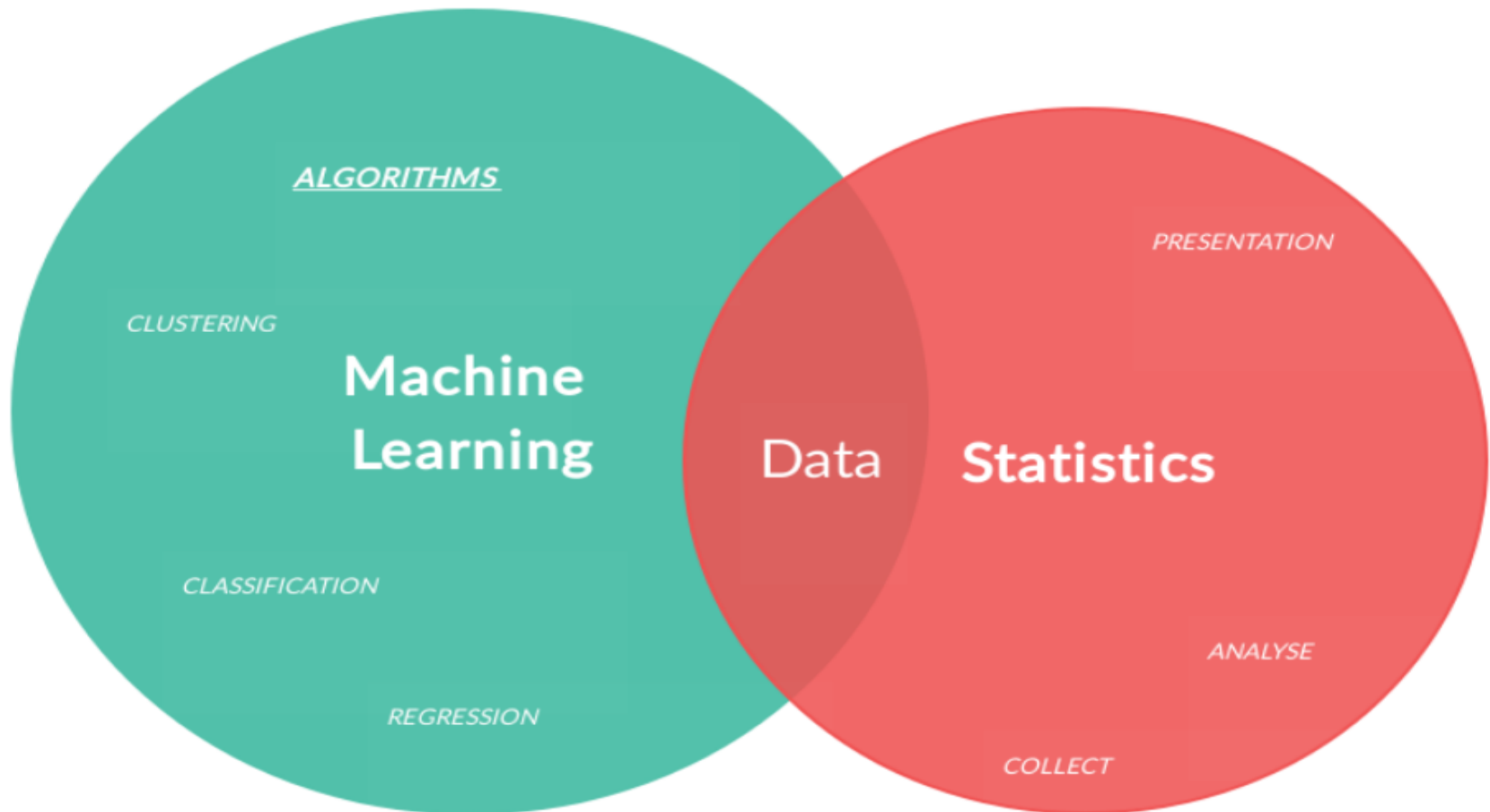


- Project Management.
- Manage stakeholder expectations.
- Maintain a Vision.
- Facilitate.
- Evangelize.

Process Owner



Statistics and Machine Learning



I WANT ANSWERS



**CALL THE DATA SCIENTIST.
RIGHT NOW!**