Google Developer Groups
Editable Location

Introduction to Data Science

Eng Ahmed ElSany

Data Science Laed @ GDG On Campus Zagazig

DATA
SCIENCE

About us

Ahmed ElSany

Data Science Lead

Osama Hassan

Data Science Co Lead

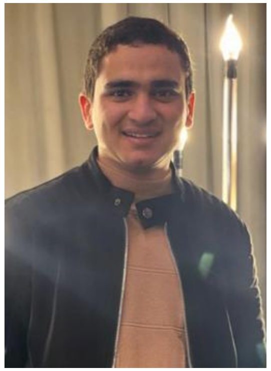
Core Team:

Ahmed Ezzat

Omar Mohamed

Mohamed Ehab

Ahmed Samir



Ahmed Hamdi

Yousef Ibrahim

Google Developer Groups

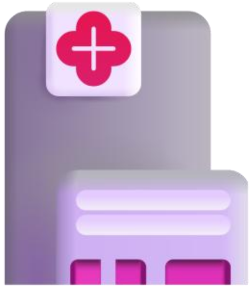
What is Data Science?

Data science is an interdisciplinary field that uses scientific methods, algorithms, and systems to extract knowledge and insights from structured and unstructured data.

Applications of Data Science

Recommendation Systems

E-commerce & Marketing

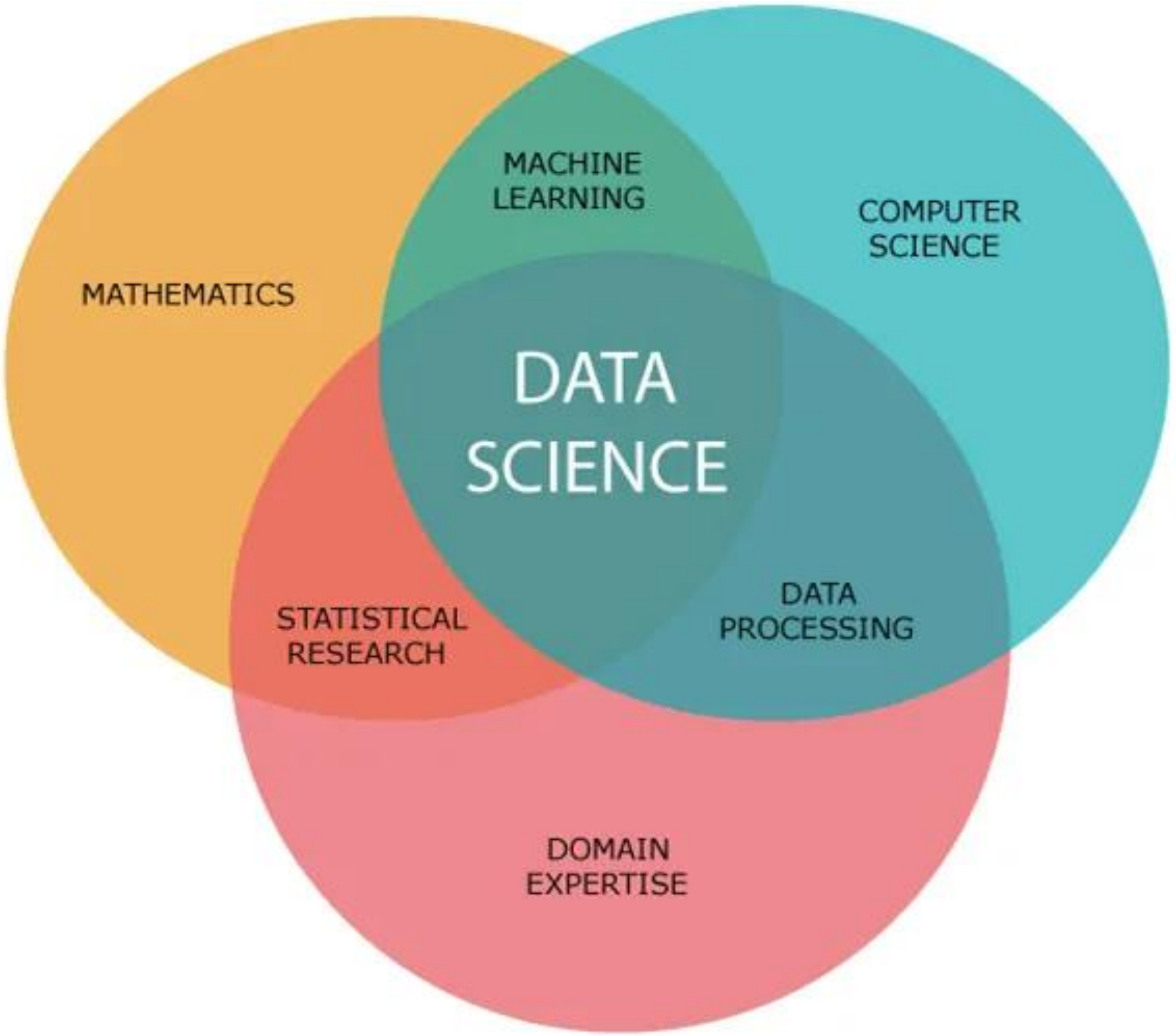


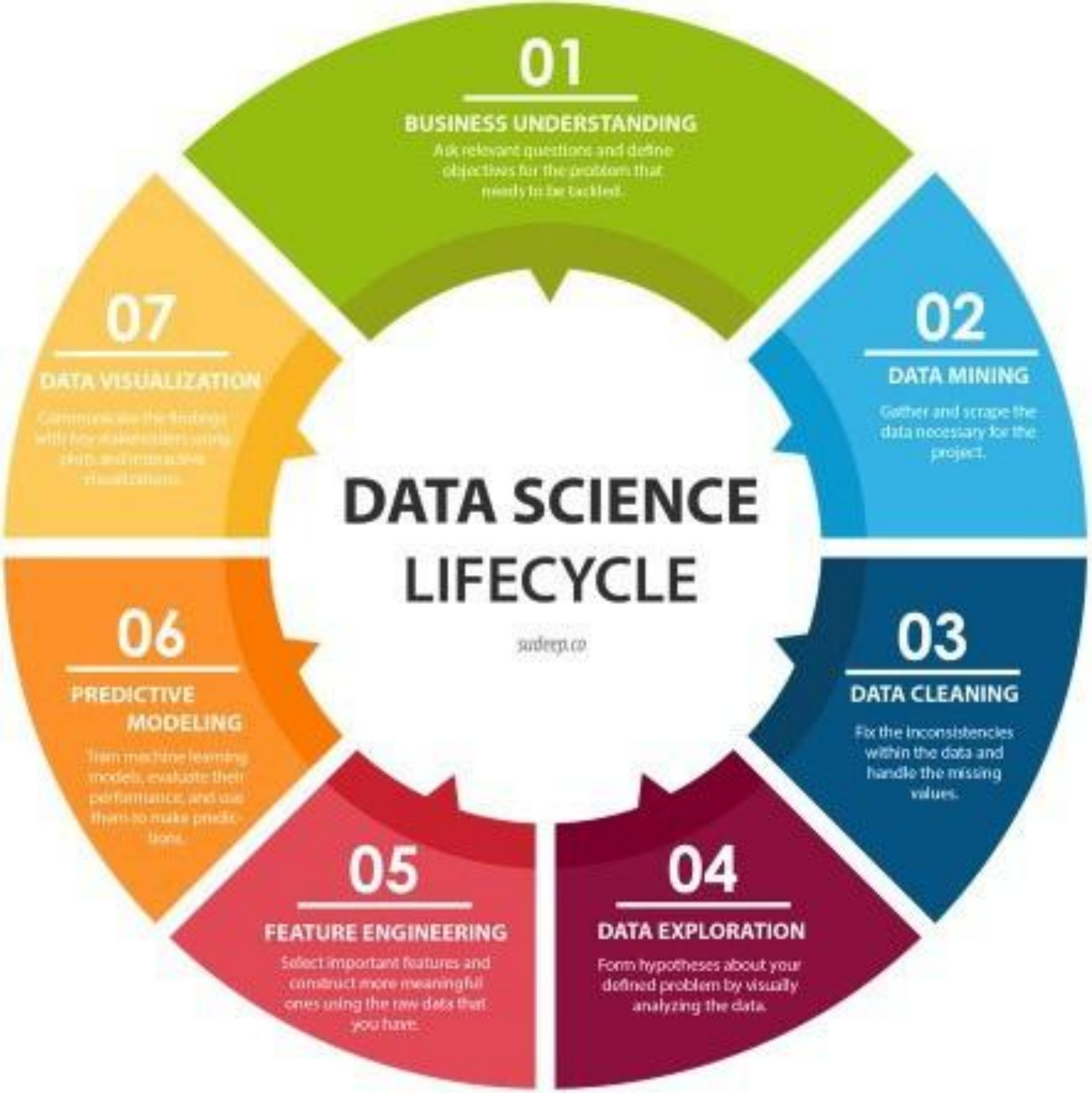
Healthcare & Medical Research

Social Media & Web Analytics

Finance & Banking

The Key Components of Data Science


MACHINE LEARNING
COMPUTER SCIENCE
MATHEMATICS
DATA SCIENCE
STATISTICAL RESEARCH
DATA PROCESSING
DOMAIN EXPERTISE



DATA SCIENCE LIFECYCLE

sudeep.co

01

BUSINESS UNDERSTANDING

Ask relevant questions and define objectives for the problem that needs to be tackled.

02

DATA MINING

Gather and scrape the data necessary for the project.

03

DATA CLEANING

Fix the inconsistencies within the data and handle the missing values.

04

DATA EXPLORATION

Form hypotheses about your defined problem by visually analyzing the data.

05

FEATURE ENGINEERING

Select important features and construct more meaningful ones using the raw data that you have.

06

PREDICTIVE MODELING

Train machine learning models, evaluate their performance, and use them to make predictions.

07

DATA VISUALIZATION

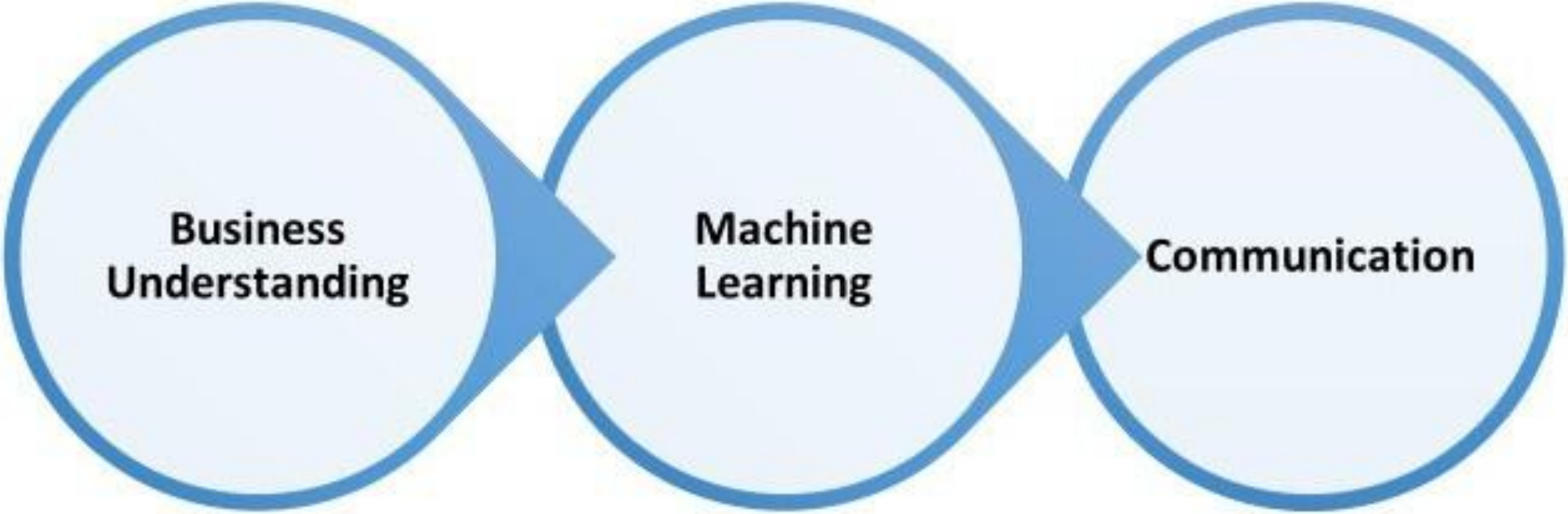
Communicate the findings with key stakeholders using plots and interactive visualizations.

Key Steps in Data Science

1. **Data Collection**
   Gathering raw data from various sources, such as databases, APIs, and web scraping.
2. **Data Cleaning and Preprocessing**
   Transforming raw data into a usable format by handling missing values, outliers, and inconsistencies.
3. **Exploratory Data Analysis**
   Analyzing data patterns, identifying trends, and gaining insights through visualization and statistical techniques.
4. **Feature Engineering**
   Creating new features from existing data to improve the accuracy and performance of machine learning models.
5. **Model Selection and Training**
   Choosing the appropriate machine learning model and training it on the prepared data.
6. **Model Evaluation and Optimization**
   Evaluating the model's performance, identifying areas for improvement, and fine-tuning the model parameters.
7. **Deployment and Monitoring**
   Deploying the trained model into production, ensuring its stability, and continuously monitoring its performance.

Google Developer Groups

Data Science Process

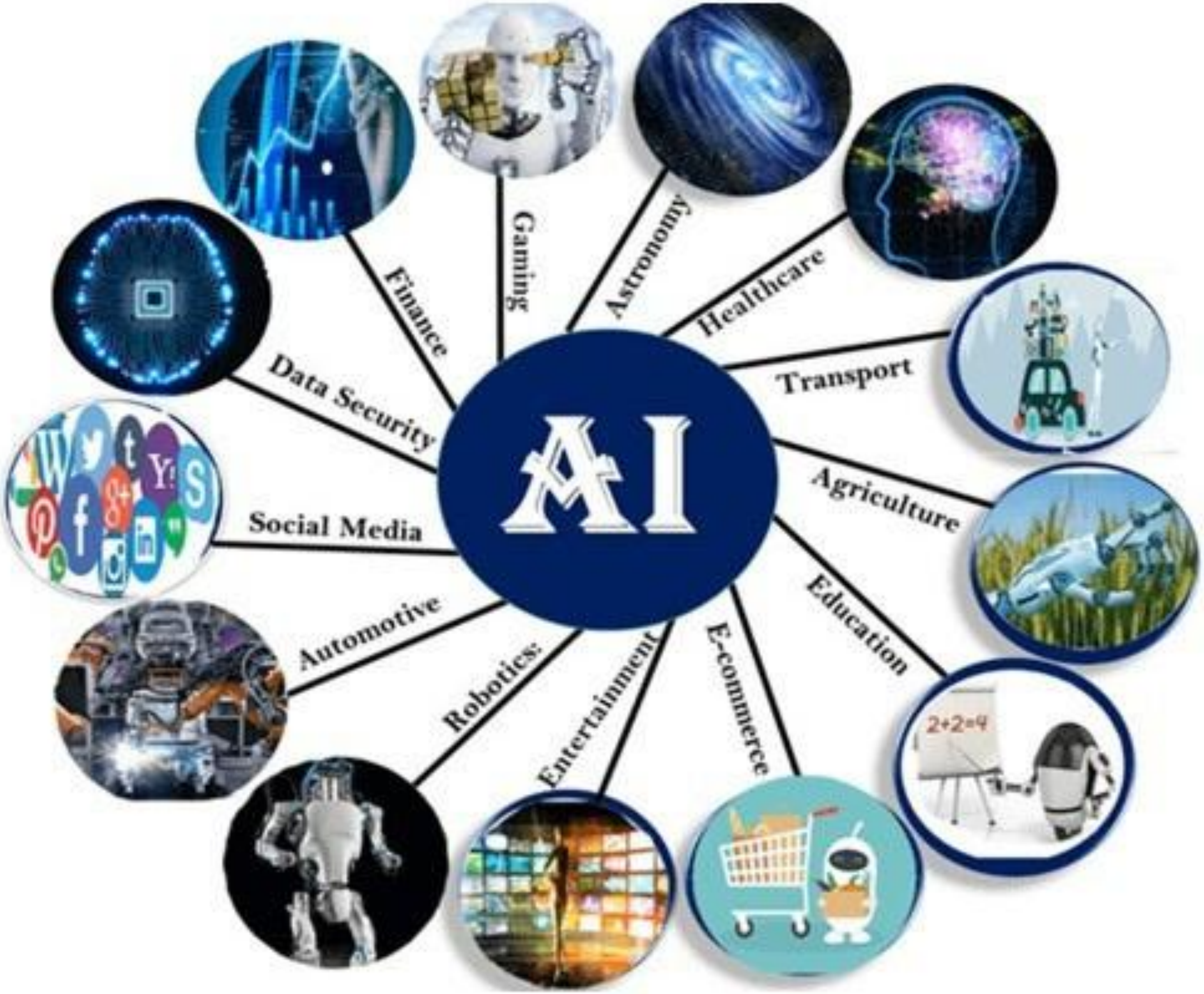


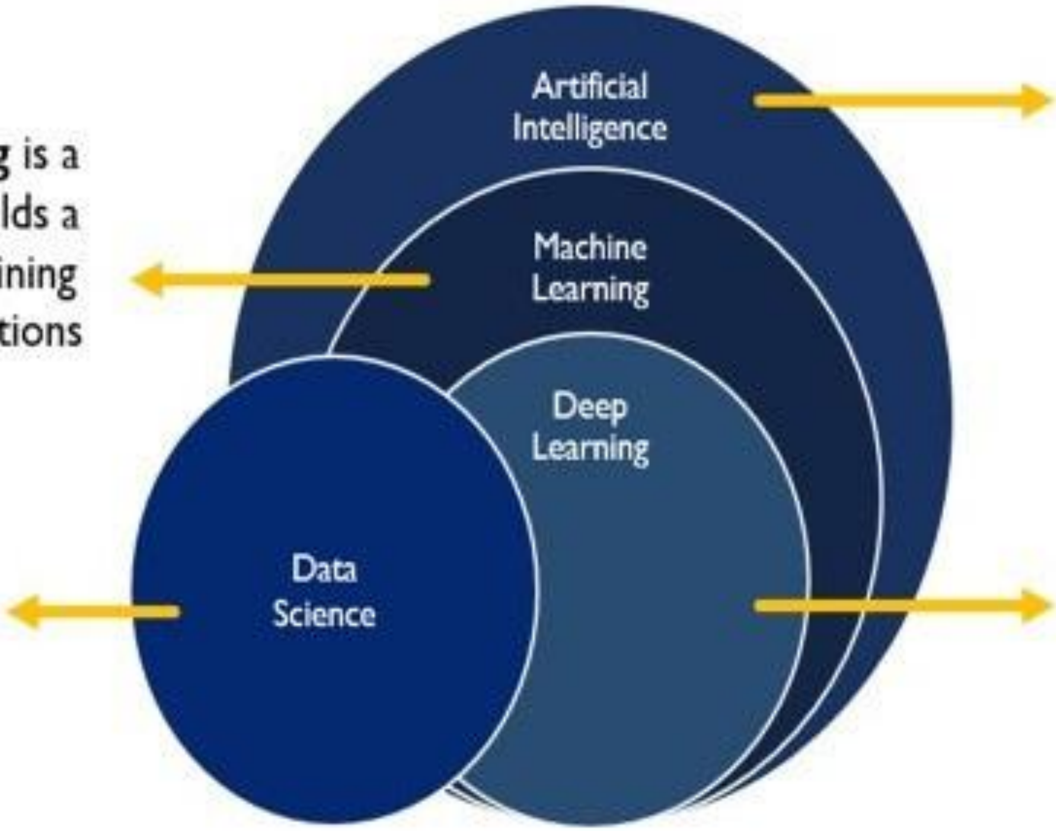
Business Understanding

Machine Learning

Communication

What is AI?


AI
Gaming
Astronomy
Healthcare
Transport
Agriculture
Education
E-commerce
Entertainment
Robotics
Automotive
Social Media
Data Security
Finance
2+2=4


Artificial Intelligence
Machine Learning
Deep Learning
Data Science

Artificial Intelligence means creating smart machines to mimic human behavior

Machine Learning is a subset of AI, and builds a model based on training data to make predictions

Deep learning is a subset of ML, a class of ML algorithms to solve complex problems.

Data Science is a subset of AI. It is an area of statistics, scientific methods, etc. to extract meaning and insights from data..

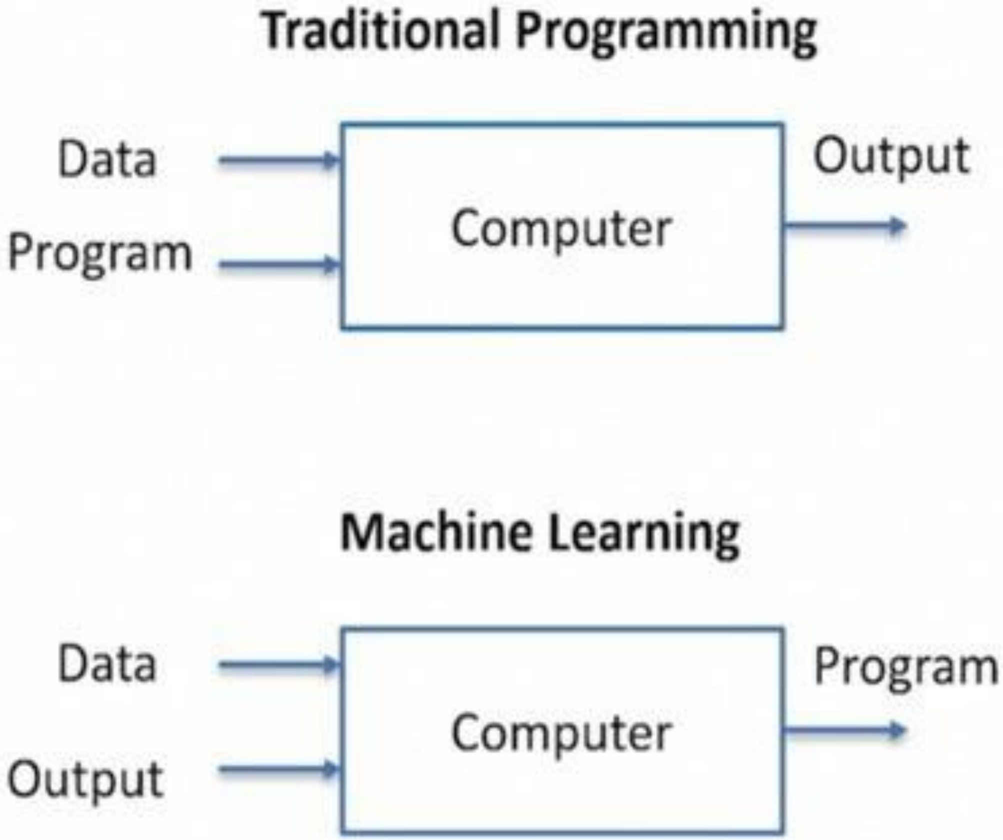
©STUDYOPEDIA All rights reserved

Comparison Between AI and Data Science

Aspect	Artificial Intelligence (AI)	Data Science
📌 **Definition**	Developing intelligent systems that mimic human cognition	Extracting knowledge from data for business insights
🎯 **Goal**	Automate processes and make smart decisions	Analyze data to find patterns and make predictions
📚 **Subfields**	• Machine Learning (ML) • Deep Learning (DL) • NLP , LLM	• Data Analysis • Data Mining • Statistics & Probability • Data Visualization
🤖 **Applications**	🚗 Self-driving cars 🗣 Virtual assistants (Siri, Alexa) 📷 Image recognition	📈 Business analytics 🔍 Movie recommendations (Netflix) 🏛 Financial analysis 📊 Big data analytics
🔄 **Relationship**	AI relies on data but uses algorithms to make intelligent decisions	Data Science can use AI as a tool for data analysis

What Is Machine Learning?

Software "models" are trained from data to predict certain patterns, trends, and outcomes

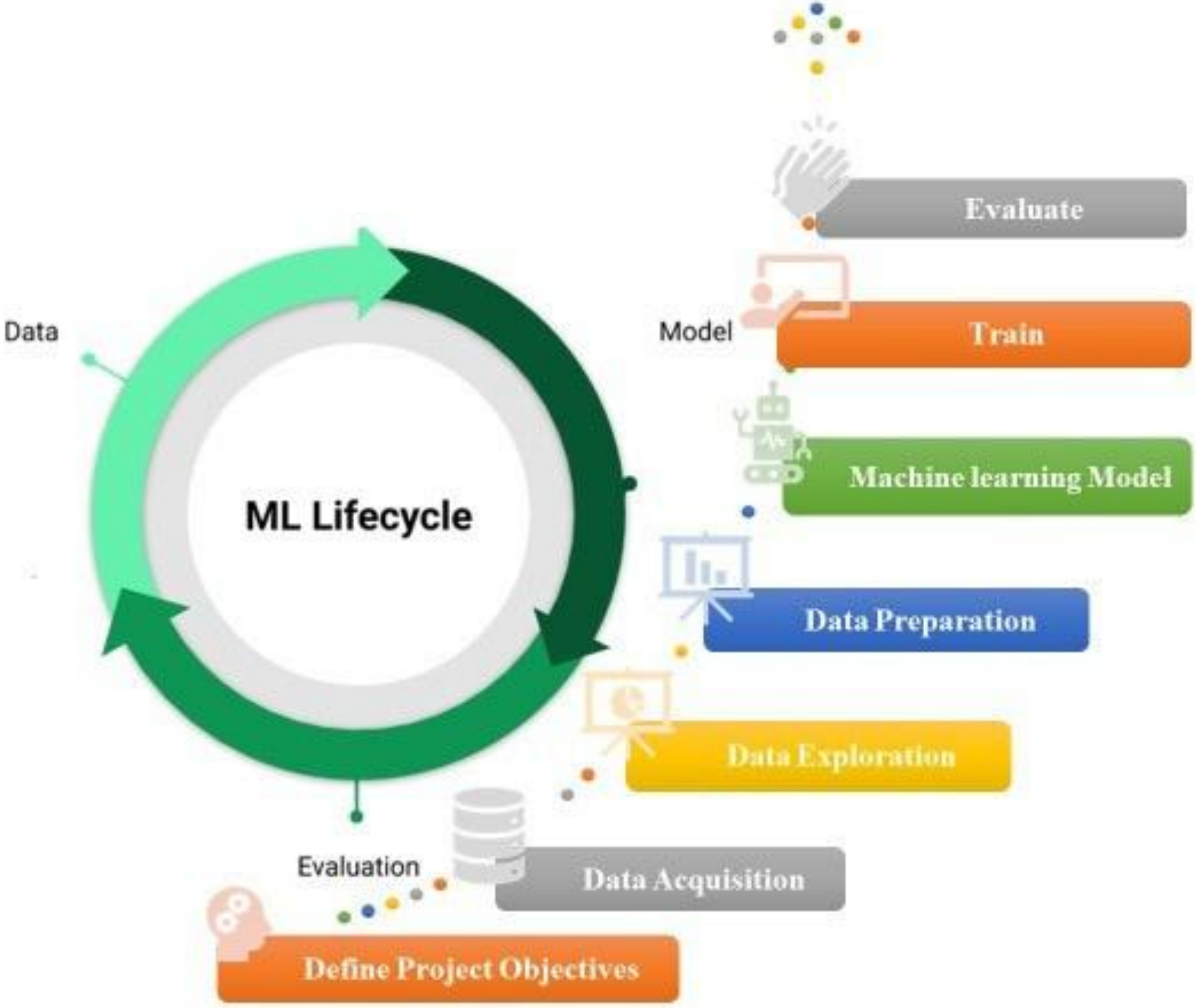


Traditional Programming

Data
Program
Computer
Output

Machine Learning

Data
Output
Computer
Program



ML Lifecycle

Data

Model

Evaluation

Evaluate

Train

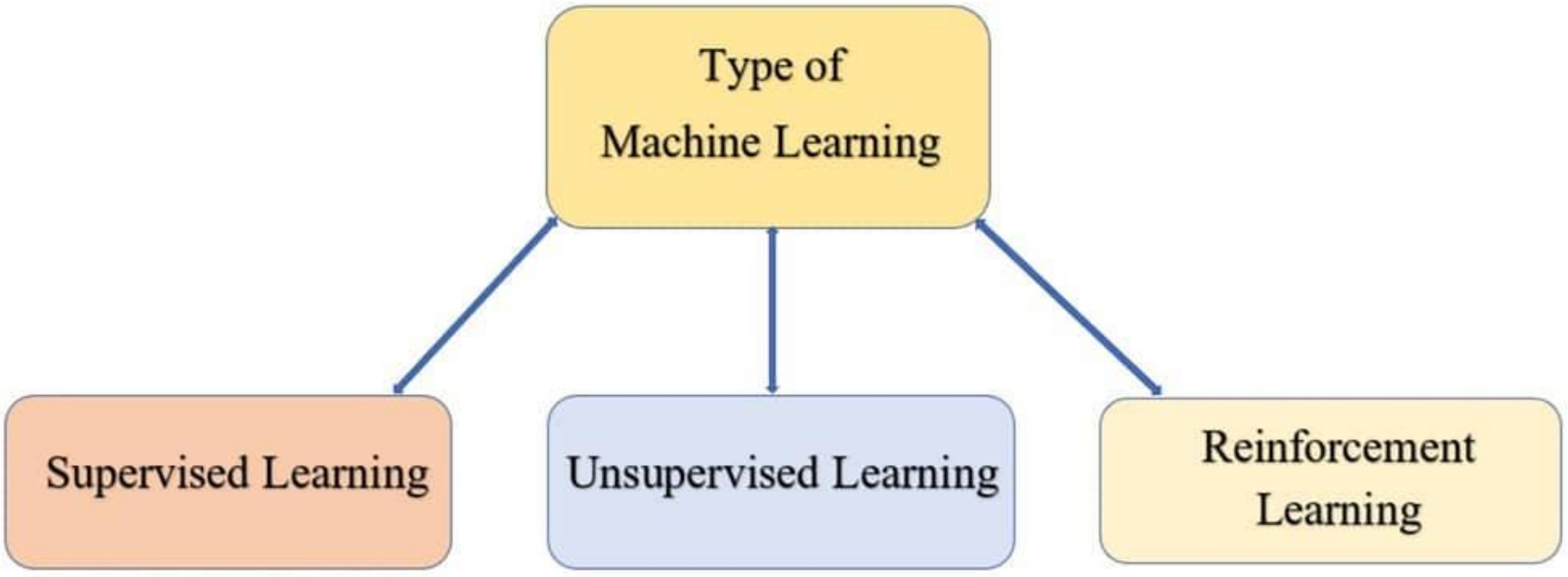
Machine learning Model

Data Preparation

Data Exploration

Data Acquisition

Define Project Objectives

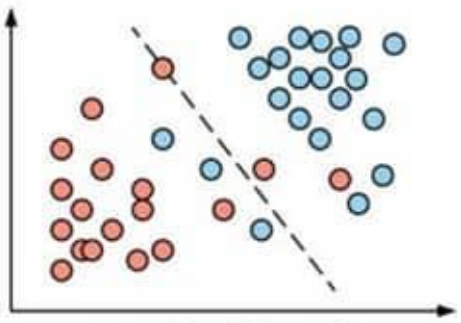


Type of Machine Learning

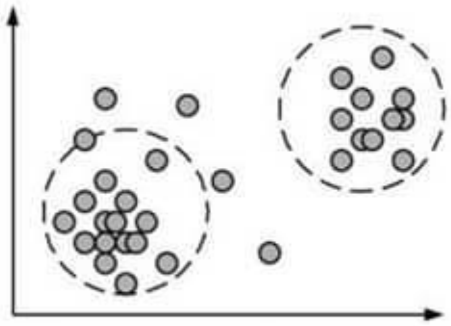
Supervised Learning

Unsupervised Learning

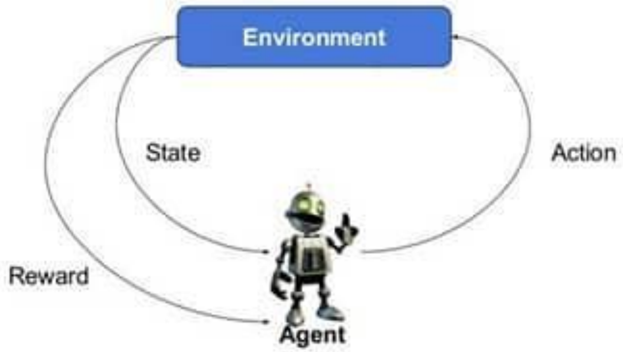
Reinforcement Learning



Supervised learning



Unsupervised learning

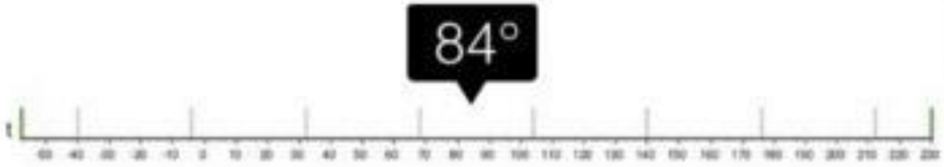

Environment
State
Action
Reward
Agent

Supervised Learning

enjoy algorithms

Regression

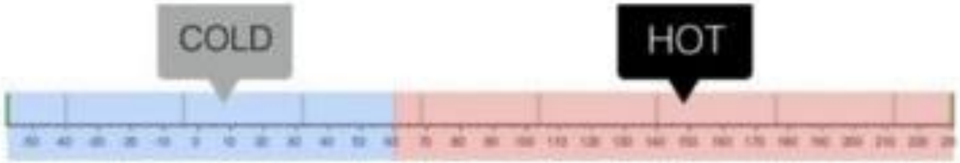
What will be the temperature tomorrow?


84°

Fahrenheit

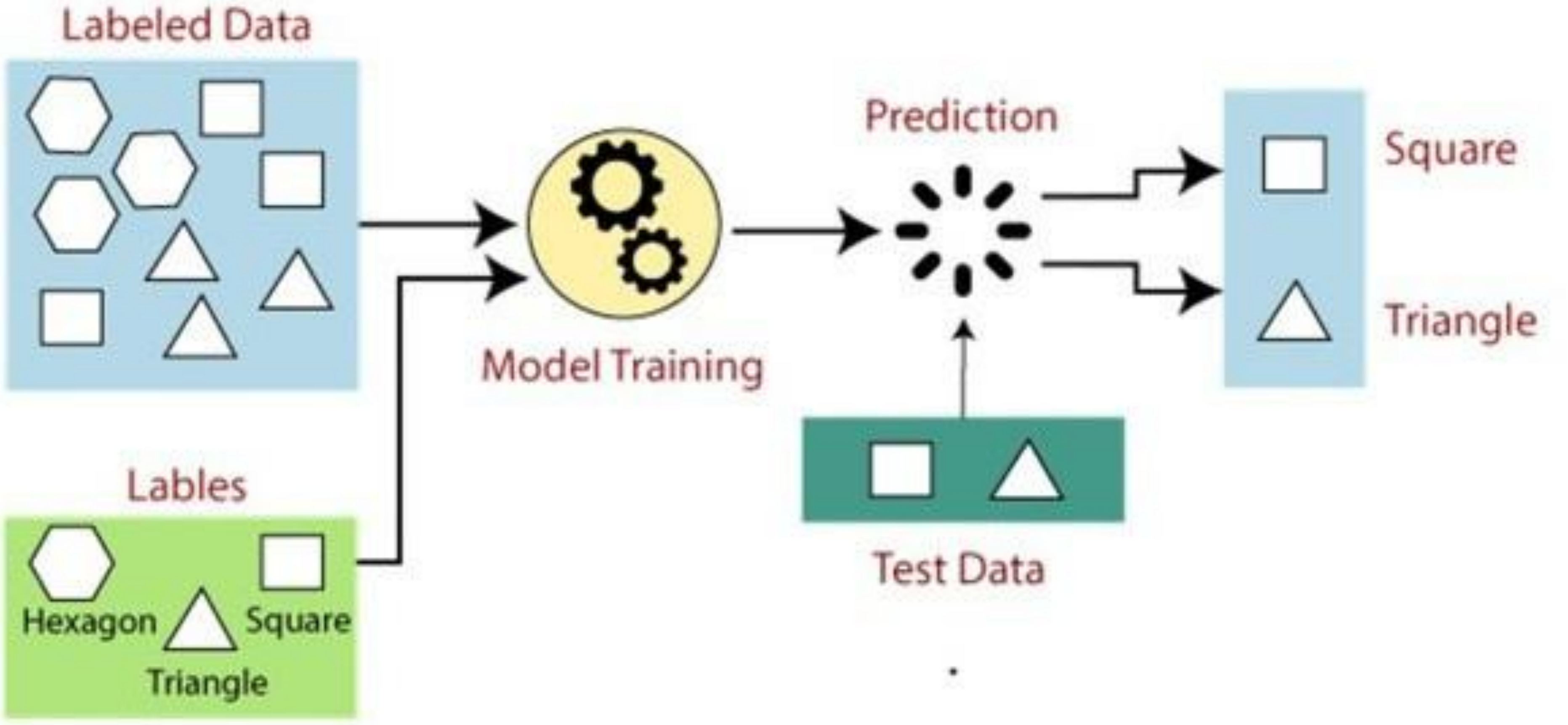
Classification

Will it be hot or cold tomorrow?

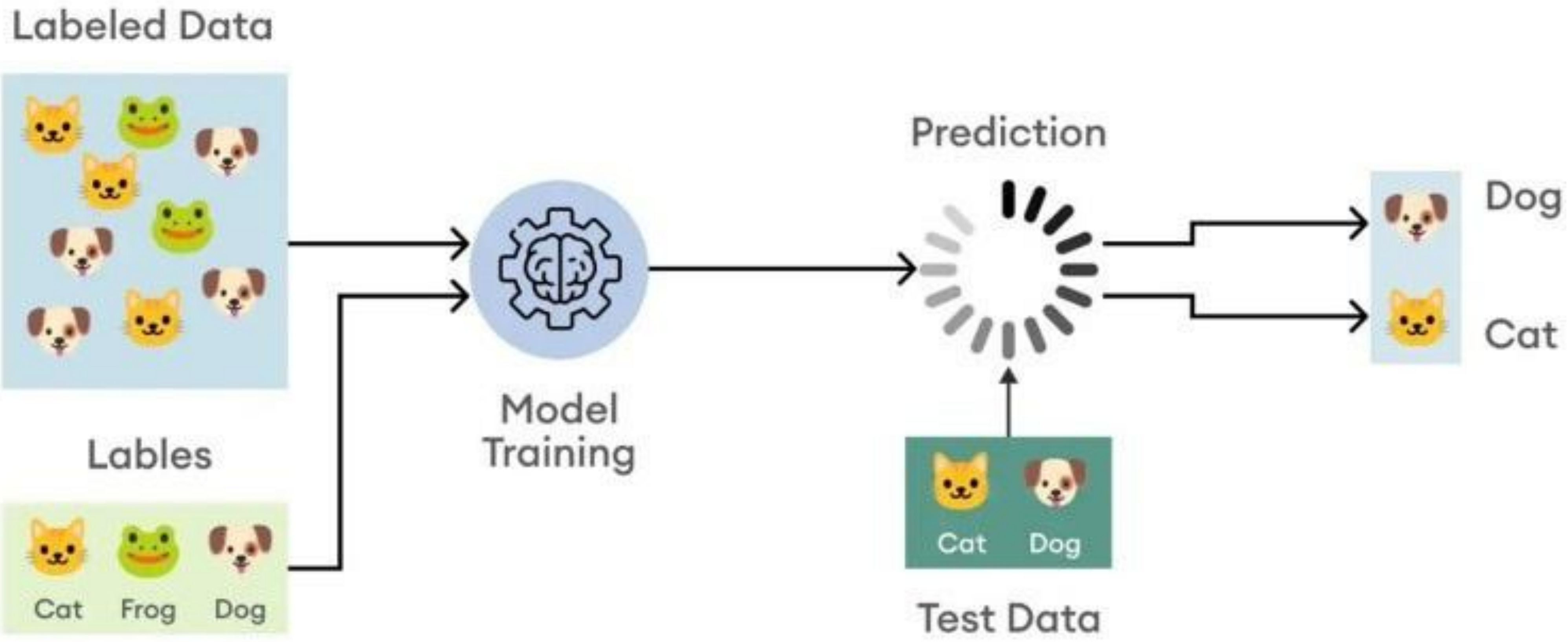

COLD
HOT

Fahrenheit

Supervised Learning

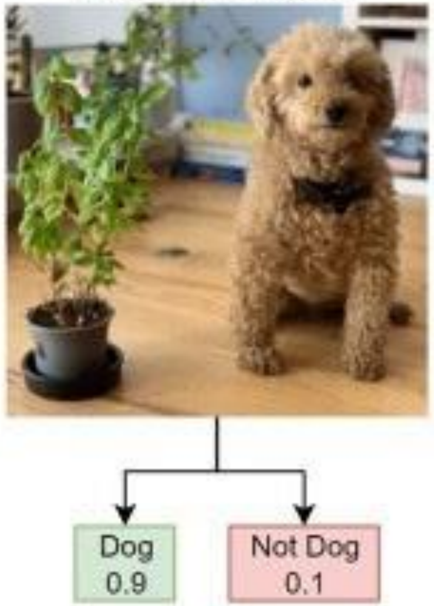

Labeled Data
Model Training
Prediction
Square
Triangle
Lables
Hexagon
Square
Triangle
Test Data

Classification


Labeled Data
Lables
Cat
Frog
Dog
Model
Training
Prediction
Test Data
Cat
Dog
Dog
Cat

Types of Classification

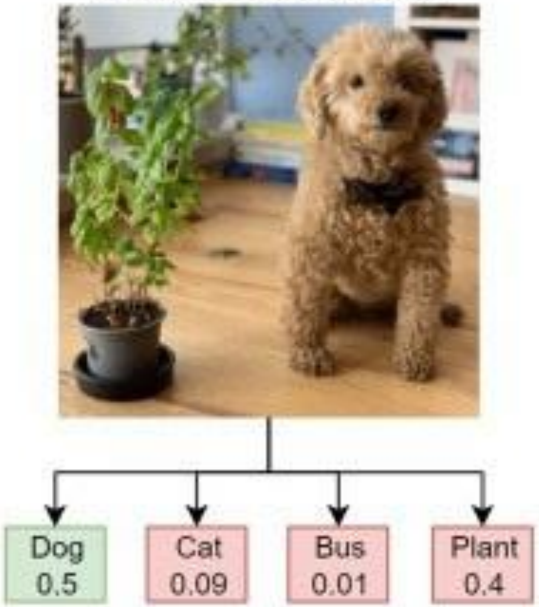
Binary Classification



Dog 0.9

Not Dog 0.1

Multiclass Classification



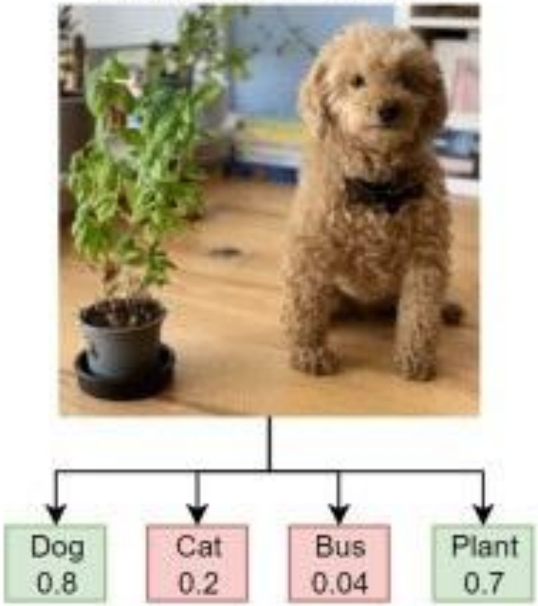
Dog 0.5

Cat 0.09

Bus 0.01

Plant 0.4

Multilabel Classification



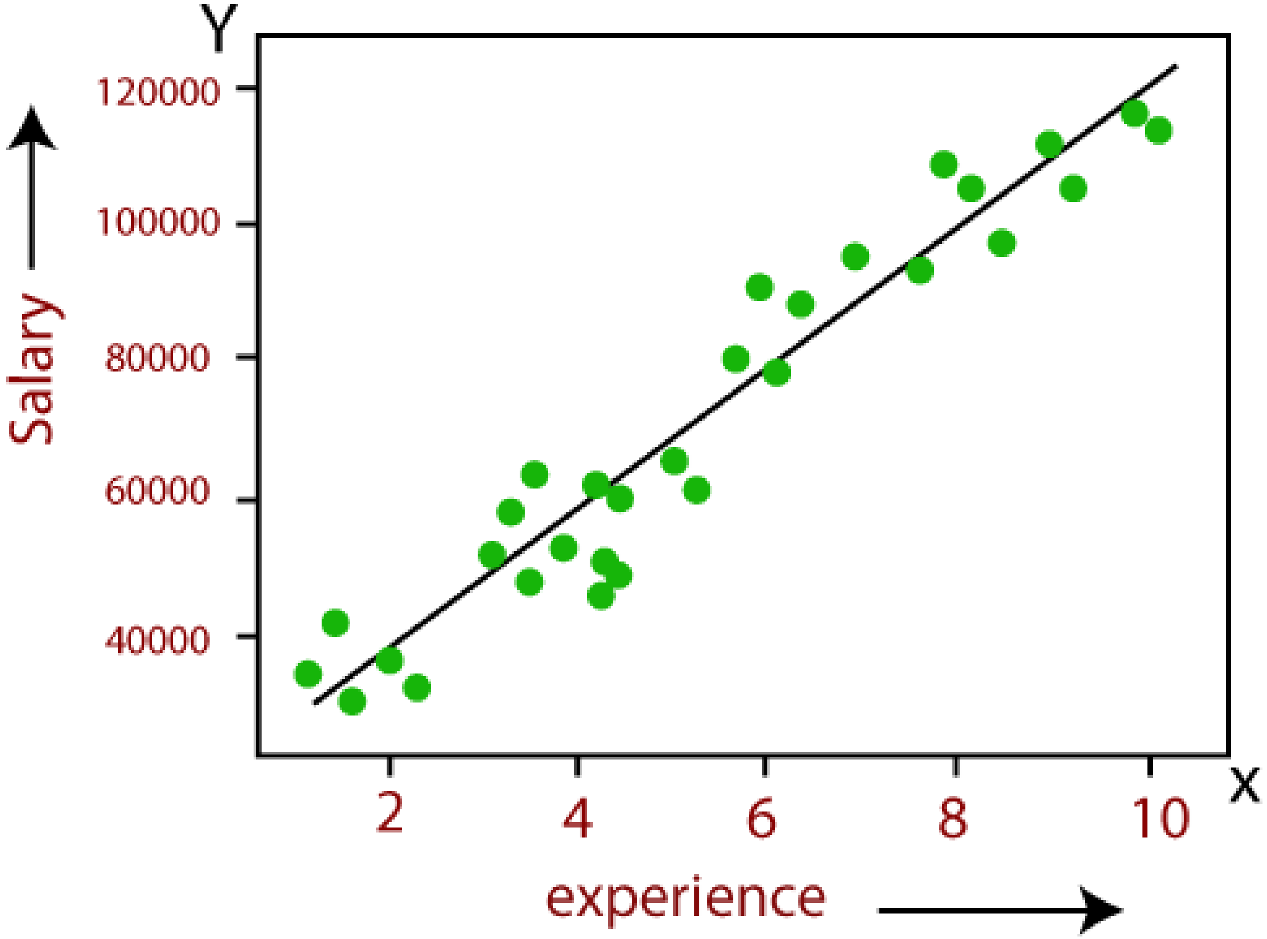
Dog 0.8

Cat 0.2

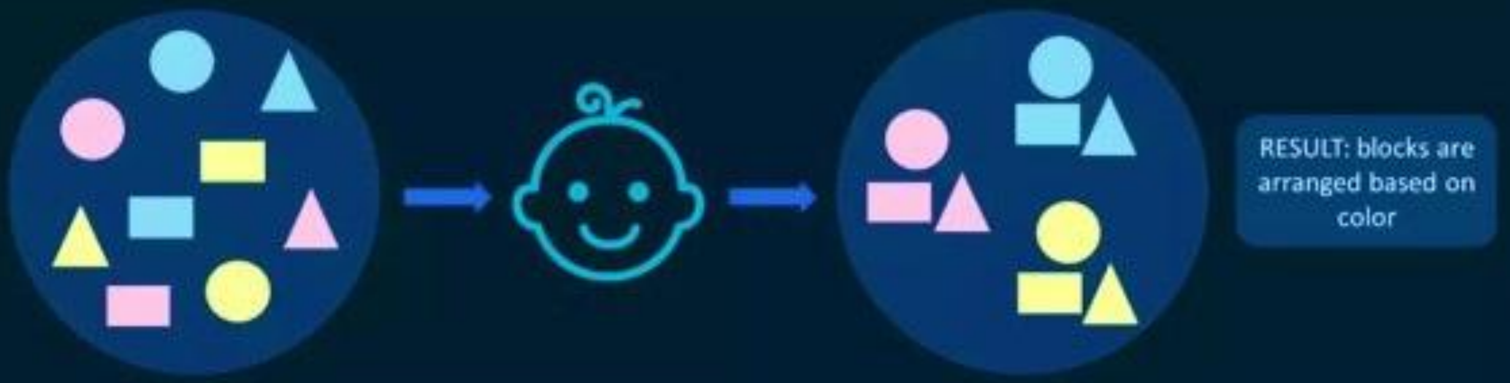
Bus 0.04

Plant 0.7

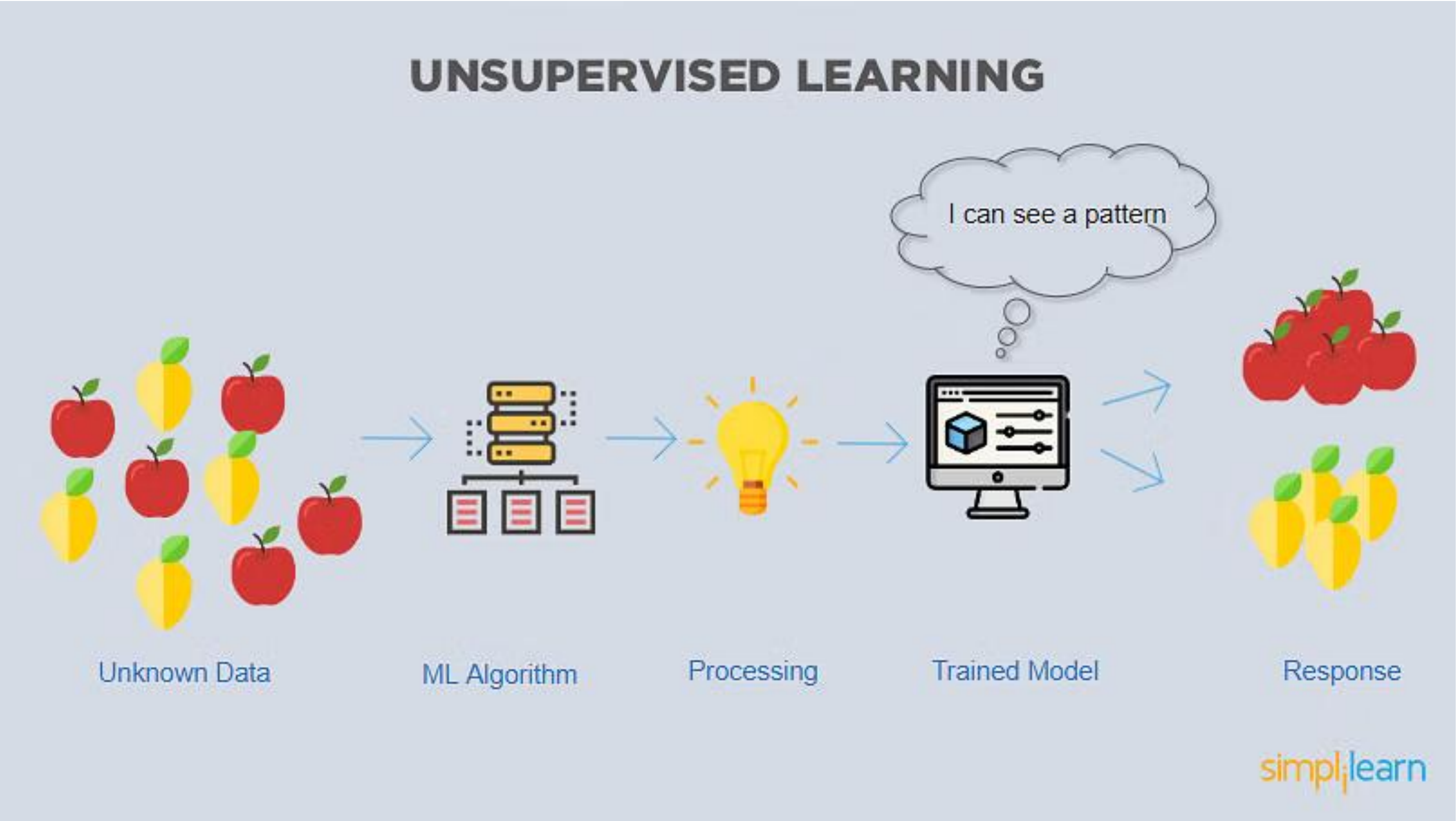
Regression


Y
120000
100000
80000
60000
40000
Salary
2
4
6
8
10
X
experience

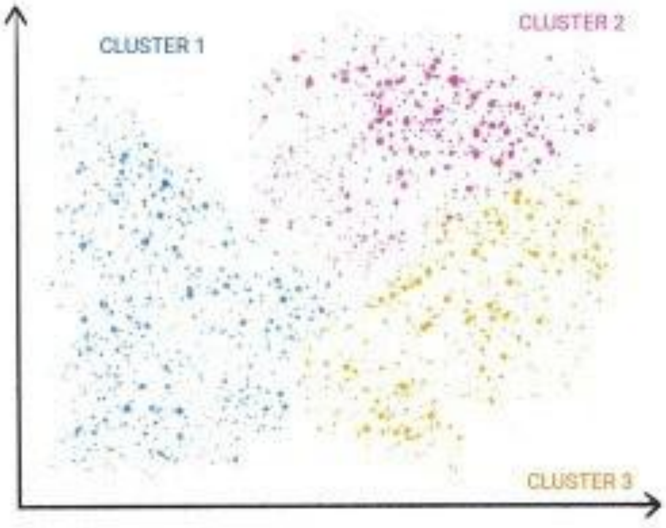
Unsupervised Learning

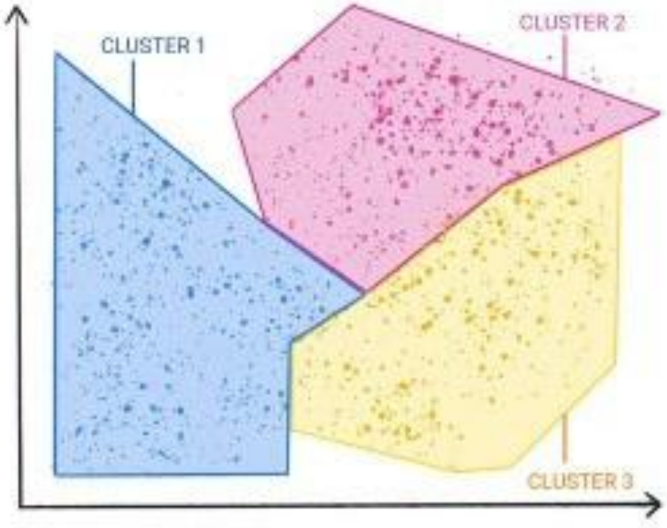


RESULT: blocks are arranged based on color

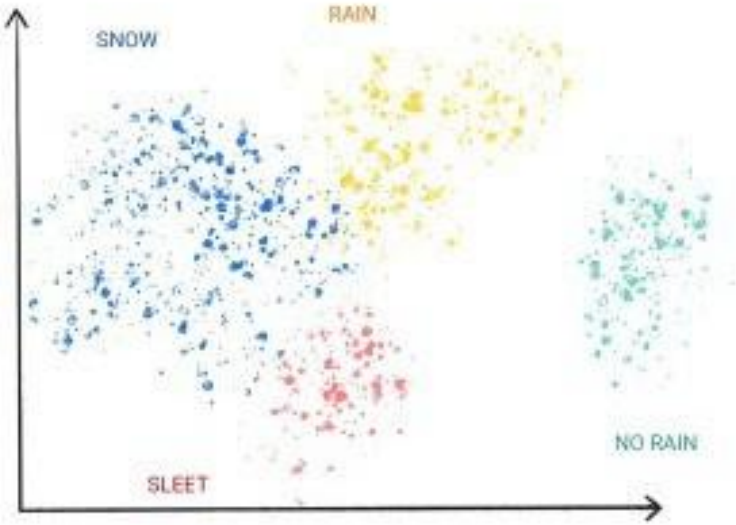

UNSUPERVISED LEARNING
I can see a pattern
Unknown Data
ML Algorithm
Processing
Trained Model
Response
simplilearn

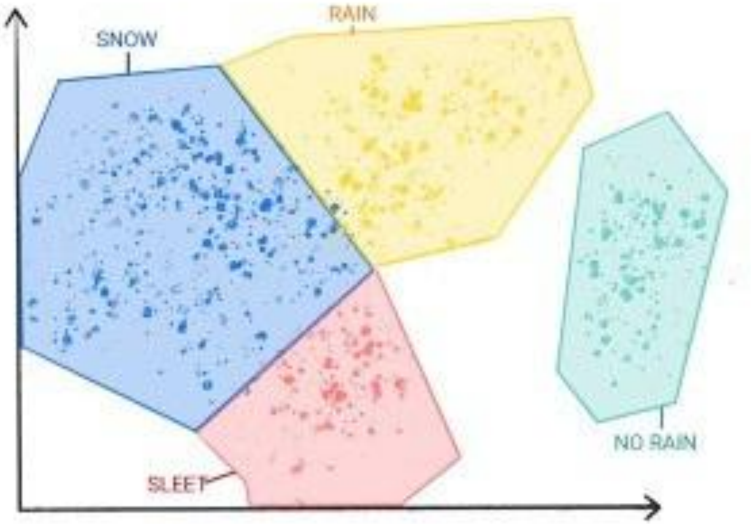
Unsupervised learning

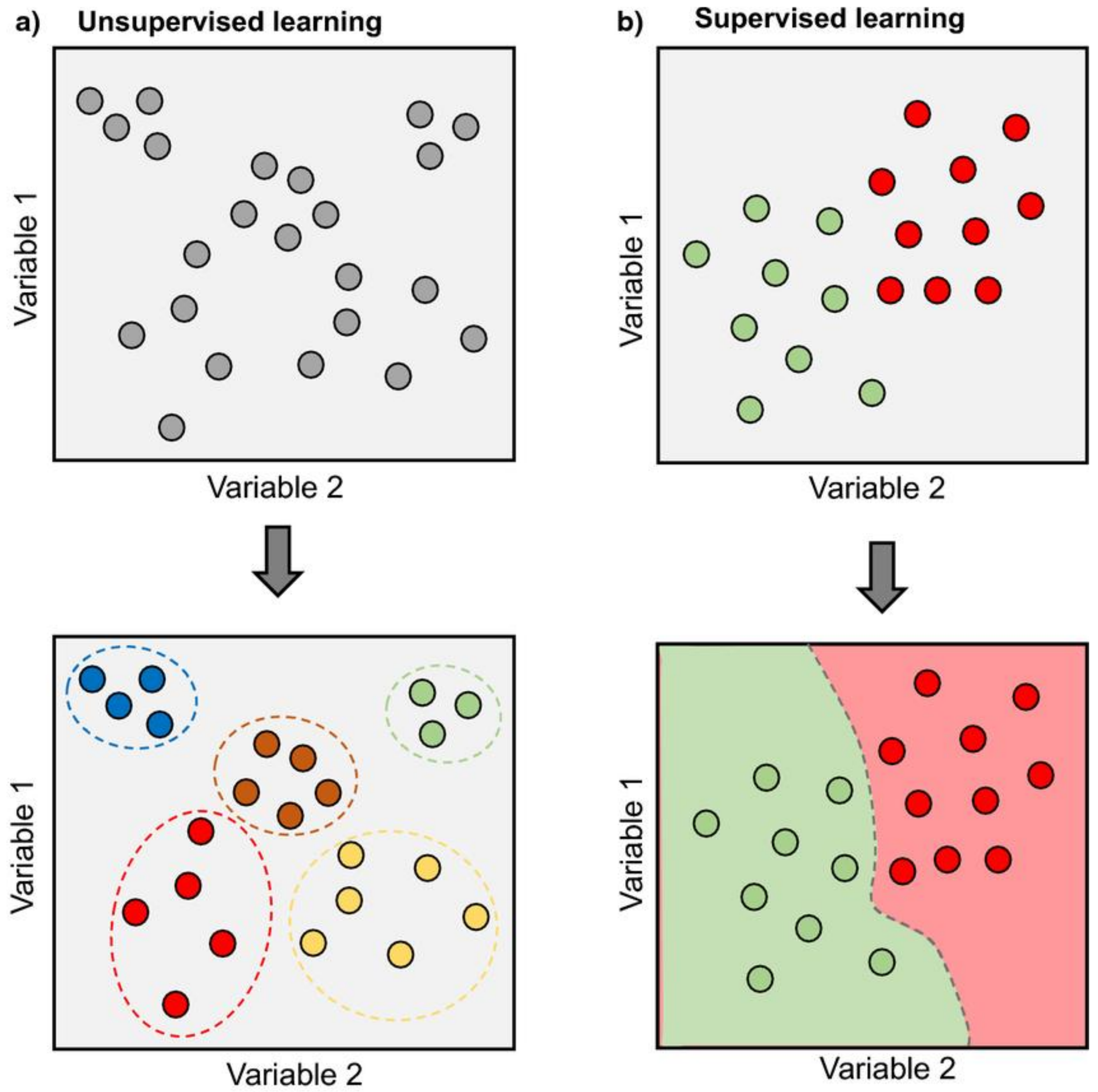




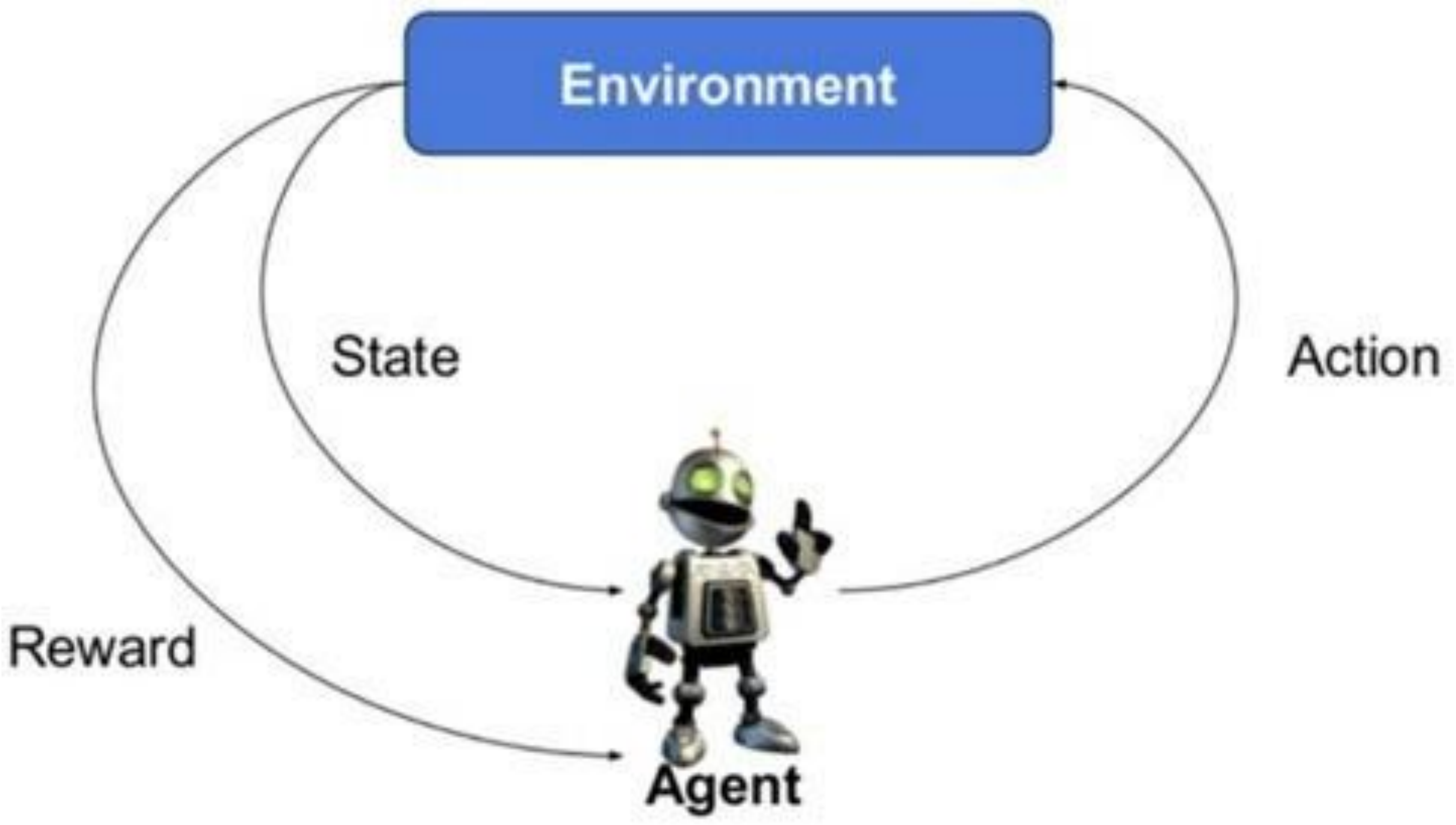
Unsupervised learning



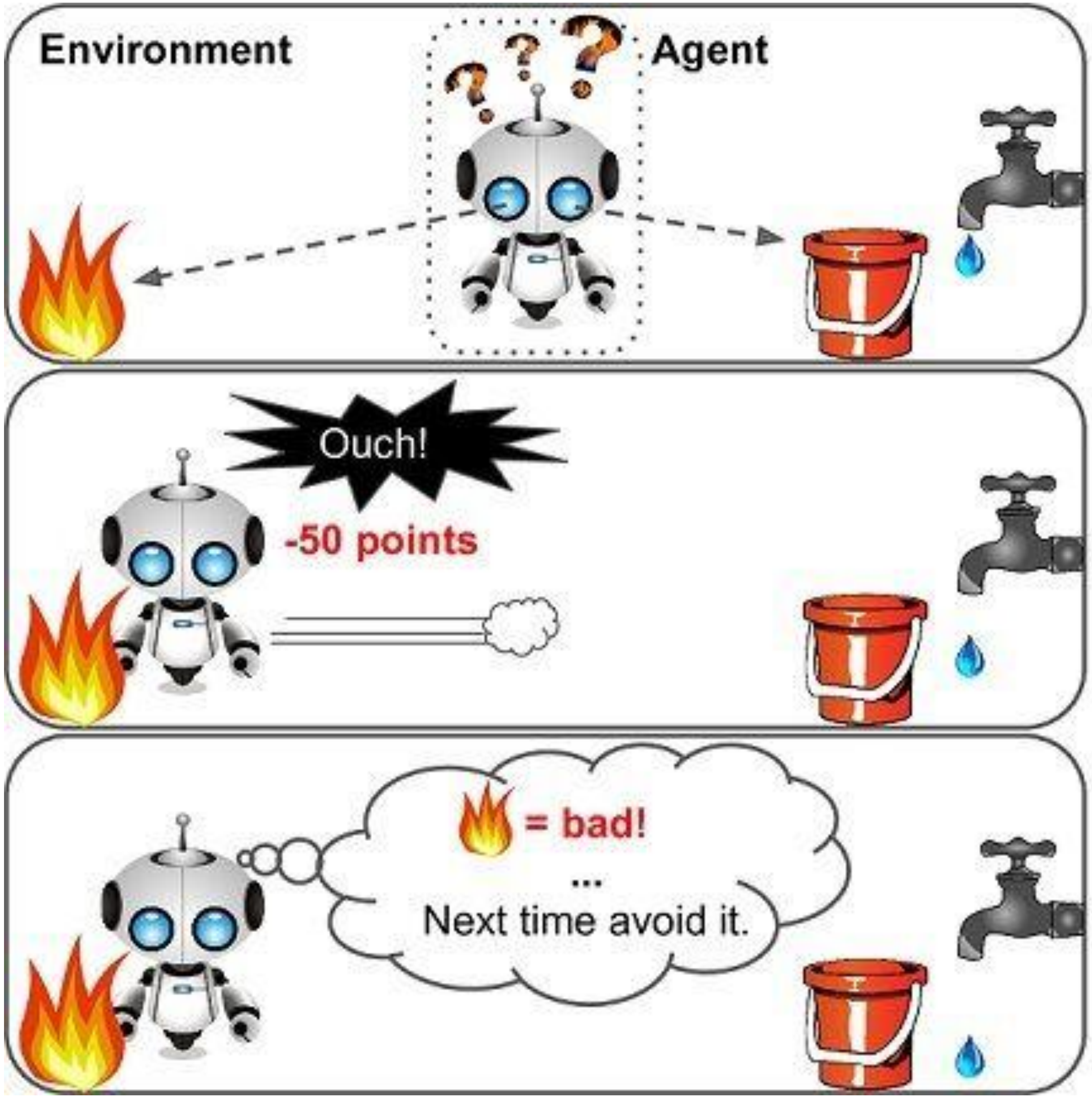



a) Unsupervised learning
Variable 1
Variable 2
Variable 1
Variable 2
b) Supervised learning
Variable 1
Variable 2
Variable 1
Variable 2

Reinforcement learning



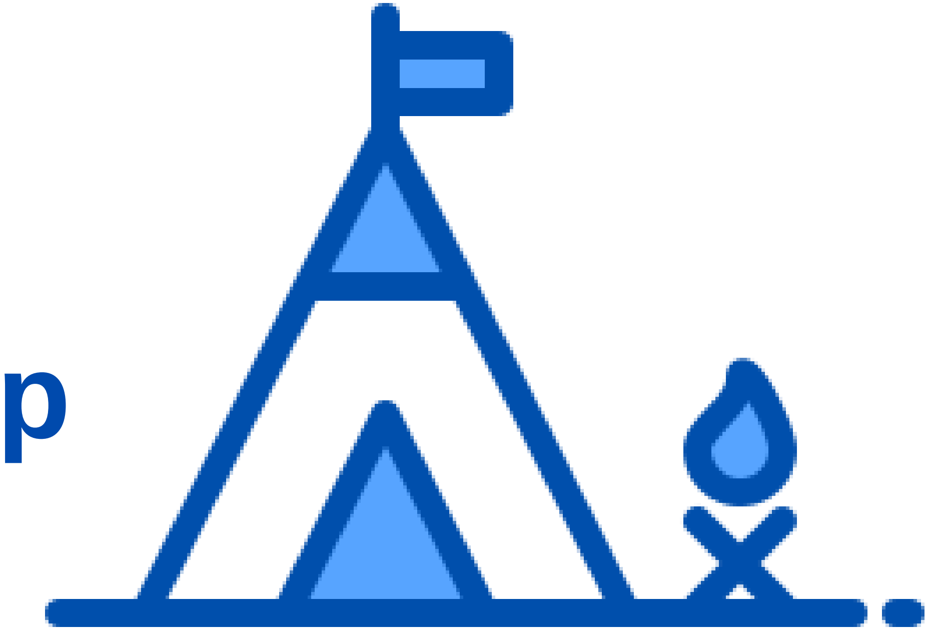
Environment
State
Action
Reward
Agent


Environment
Agent
Ouch!
-50 points
= bad!
...
Next time avoid it.

1. Observe
2. Select action using policy
3. Action!
4. Get reward or penalty
5. Update policy (learning step)
6. Iterate until an optimal policy is found

Google Developer Groups

TimeLine Bootcamp



TimeLine Bootcamp

DAY 1

INTRO AI
Statistic&Math

DAY 2

Python
Basics

DAY 3

Pandas
Numpy

DAY 4

Seaborn
Matplot

DAY 5

Project
Data Analysis

DAY 6

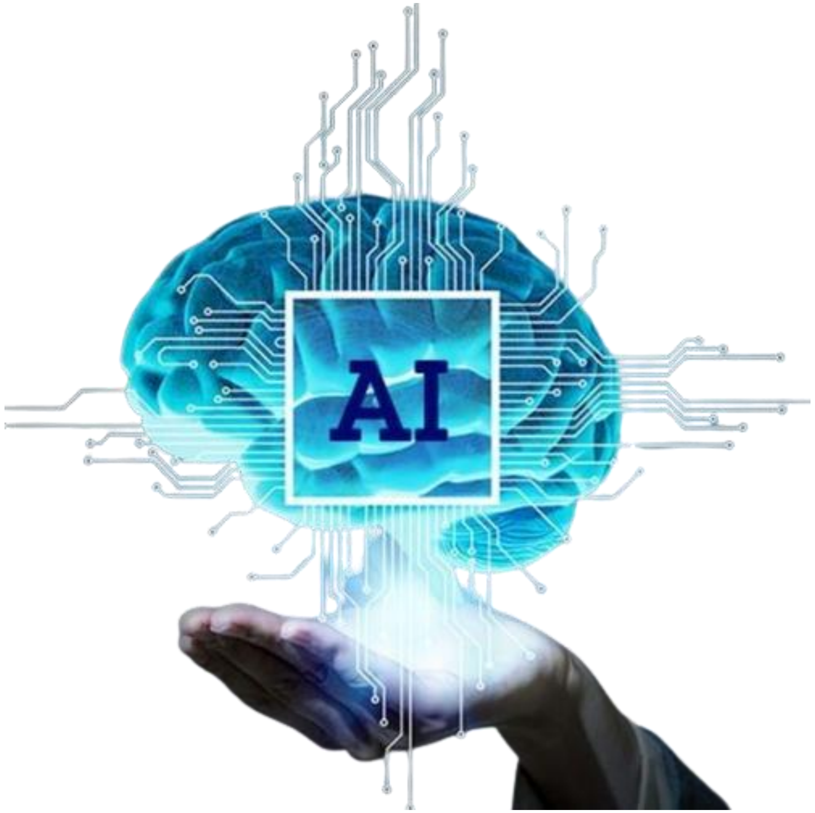
ML

DAY 7

ML
Online

Google Developer Groups

Thank You!



Google Developer Groups