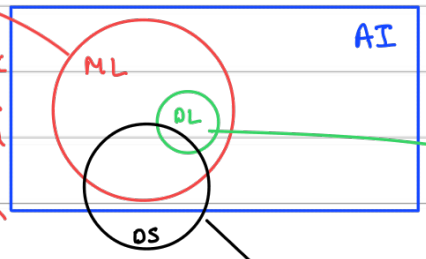
AI vs ML vs DL vs DS

AI: Smart applications that can perform its own task without any human intervention.
Ex: Self driving cars
Robots



ML is a subset of AI
It provides stats tools to learn, analyse and develop predictive models from the data
Ex: Amazon, Netflix ⟹ Recommendation System
Classification Apps
Regression Application

Data Science:-
covers all DL, ML, AI

Deep Learning (DL) used to mimic Human Brain. done through "Multilayered Neural Network".
Ex: Object Detection
Image Recognition
Chat bots
Chat GPT (from transformers)

Supervised vs Unsupervised Learning

Types of Machine Learning:-

1) Supervised ML
2) Unsupervised ML
3) Semi supervised ML
4) Reinforcement Learning

① Supervised ML:-

- Used to solve 2 type of problem statements:-
  - a → Classification
  - b → Regression problem

- In Supervised ML, dataset has a specific o/p and Model is trained to predict this o/p.

Ex:- Classification

Independent Features

Play hours	Study hours	Pass/Fail
8	2	Fail
7	3	Fail
3	5	Pass

dependent Feature (depends on Play & Study hours)

as 2 categories ∴ it is called as "Binary Classification Dataset"

The above data o/p is used to train our Binary Classification model.

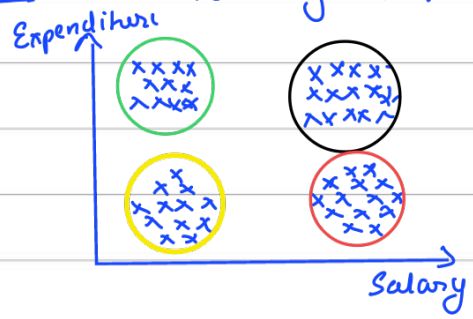
Ex: Regression:- o/p → continous

Size of House No of Rooms Price of House (continous)

② Unsupervised ML :-

- Only dataset (without any o/p) is available and Clusters and similar groups are formed of the data.

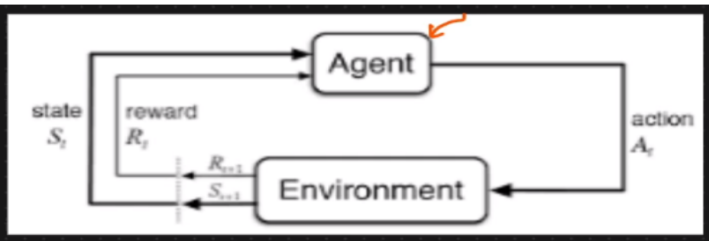
Ex: Customer Segmentation.



Clusters formed from a shopping mall data.
Now when a new product is launched then this clusters are used to target the audience with highest expenditure.

③ Semi-Supervised ML :- [Supervised + Unsupervised]

④ Reinforced Learning:-



Reinforcement learning is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

TRAINING, TEST & VALIDATION DATASET

DATASET

Befor model training and prediction the data set is divided into

- **TRAINING DATASET**
  - model trained using this dataset
- **VALIDATION DATASET**
  - "Hyper Parameter tuning" performed using this dataset
  - # use to tweek some parameters of model to increase accuracy
- **TEST DATASET**
  - used to test model accuracy
  - # Models will not know anything about the test data

① MODEL PERFORMANCE :- accuracy ↑↑ → High

OVERFITTING & UNDERFITTING

Ex 1) Dataset

(ex: Book) Train set → Model is trained → Accuracy High

(ex: Exam paper) Test set → Model is tested → Accuracy decreases

} Over-fitting

in overfitting condition model will perform well if question comes from the book (i.e training dataset) otherwise if the question is from Test set the accuracy will be low. i.e

5☆ Overfitting : low Bias (Train sets) and High Variance (test set)

↳ if training accuracy is low then high bias also if test accuracy is high then low variance.

Ex2:

Train → Accuracy ↓↓(low) → 55%
Test → Accuracy ↓↓(low) → 50%
} Underfitting

$\boxed{\text{Underfitting : High bias \& High Variance}}$

Ex3:

Train → accuracy High → 85%
Test → accuracy High → 86%
} Generalised Model

$\boxed{\text{Generalised Model : low bias \& Low Variance}}$

↳ ☆ our aim as a data scientist