MACHINE LEARNING (ML)

- ML is a type of AI that provides comps. with the ability to learn without being explicitly programmed.
- It focuses on the development of comp. progs. that can teach themselves to grow & change when exposed to new data.

ML Process Flow :- Using the data, the system learns an algo. & then uses it to build a predictive model. The system then performs the recommended tasks & uses feedback data to tune the model to be more accurate.

Training Data → Learn Algo. → Build Model → Perform → Feedback

Applications & Use Cases of ML :- Advanced Analytics, NLP, Chatbots, ANN

- Healthcare : Patient diagnosis
- Finance : Fraud Detection
- Manufacturing : Anomaly detecⁿ
- Retail : Inventory Optimizaⁿ
- Govt : Smarter Services
- Transportaⁿ : Demand Forecasting
- N/ws : Intrusion detecⁿ
- E-commerce : Recommender Systems
- Media : Interaction & Speed
- Educaⁿ : Research Insight.

Datasets :
⇒ data.gov.in – govt. website for data regarding health & fam. welfare, travel n tourism, water n sanitaⁿ, etc.
⇒ who.int/data – during covid, a lot of models were using data from the WHO website.
⇒ UCI Machine Learning Repository
⇒ Kaggle.com

Use cases :-

① Siri – Its an intelligent ~~agent~~ personal assistant. Apple claims that the s/w adapts to the user's individual preferences over time & personalises results about which apps to use for which purpose, it plays the songs u want to hear, it gives u direcⁿ's, it wakes u up, & much more.

② Financial Services – It uses ML tech. for 2 reasons :
- to identify imp. insights in data for investment opportunities.
- to prevent fraud by using data mining to identify clients with high risk profiles.

③ Healthcare – Wearable devices & sensors can use data to assess a patient's health in real time.
- Also the tech. can help medical experts analyze large amt. of data to identify trends that may lead to improved diagnoses & treatment.

④ Retail :- Websites recommending items u might like based on previous purchases are using ML to analyze ur buying history – & promote other items u'd be interested in. Also k/a Personal Recommendaⁿ system.

⑤ Govt : A lot of govt. schemes are built around from these datasets. ML models are made to be able to assess & overcome the deficiencies shown by these models. Eg a particular area suffering from TB.

Learning Process in ML :-

Tradiⁿal Programming Paradigm

Rules → | Tradiⁿal Programming | → Answers
Data → | |

Machine Learning Paradigm

Ans/Labels → | Machine Learning | → Rules
Data → | |

Data → | Model | → Inferences

Components of Learning :-

- Collecting & Preparing data
- Choosing & Training Model (Training data)
- Evaluating a model (Test data)
- Hyperparameter Tuning & Predicⁿ
  (Parameters which we can adjust/tune by ourself to inc the efficiency of the model)

Use Case :- Help John predict the price of an apartment which depends upon various factors ~~like~~ :-

Price	Apartment Name/No.	No. of bedrooms	Floor No.	Criminal rate per year	Pollution Level	Distance from Educaⁿal Inst

① **Data Acquisiⁿ** * It involves acquiring data from all the identified internal & external sources that can help answer the business quesⁿ.
- This data could be,
  - logs from webservers
  - social media data
  - census datasets
  - data streamed from online sources via API

② **Data Preparaⁿ/Data Wrangling/Data Cleaning.**

- Data should be consistent like floor nos. should either be 0,1,2.. It should not be like somewhere ground floor is denoted by 'G' & somewhere
- No missing values should be there, either remove that record or fill it by some avg. value. (like Pollⁿ level).
- Instead of using exact values for distance of educaⁿal instituⁿs, can use binary values (T/F), like educaⁿal instituⁿ within 1km radius. It will make the model simplistic.

* Data wrangling is the process of cleaning & unifying messy & complex data sets.

* Data after reformatting can be converted to JSON, CSV or any other format that makes it easy to load into one of the data sci tools

③ Hypothesis & Modelling – Based on the requirements, a model is created using the dataset.

All atts. are independent of each other { No. of bedrooms ⟶ / Criminal Rate ⟶ / Pollu^n ⟶ } → $f(x)$ ⟶ Price

Acc^{ing} to my hypothesis, ~~ma~~ price depends upon these 3 factors only.
And this hypothesis could either be accepted or rejected.

* This process involves formatting & testing hypothesis about the data & the process that generate it.
* Requires writing, running & refining the progs. to analyze & derive meaningful business insights from data.
* Mostly written in languages like Python, R, Spark.

④ Evaluation & Interpreta^n :* The model is evaluated using test dataset.

Testdata ⟶ $\boxed{f(x)}$ ⟶ Accurate?

* If accuracy is low, the above steps are repeated until a good model is found.
* Accuracy depends upon the type of applica^n we r making the model for, like if we r making a model for predicting ~~fee~~ of students using previous grades, attendance etc., then 80–90% accuracy is sufficient. But in case of medical models, it should be very much accurate.

⑤ Deployment :* Here, the model we created is deployed into the market.

* Models generally have to be ~~decoded~~ recoded before deployment
eg. data scientists may favor Python, but produc^n envs. may require Java.

⑥ Operation & Optimiza^n :- Retraing the model using new factor from dataset.

$\boxed{f(x)}$ ⟶ $\boxed{y(x)}$ ⟶ Accurate
New factor from dataset (eg. educa^{n}al institu^{n}s) ⟶ $\boxed{y(x)}$

* After the model is retrained, we evaluate the model & deploy it.

Ingredients of ML :-

- → Tasks
- → Model
- → Features

⇒ Select best features
To design a model
To perform right set of tasks

Types of ML
- → supervised learning
- → Unsupervised Learning
- → Semi-supervised Learning
- → Reinforcement Learning.

Supervised Learning - In this type of learning, the dataset on which the machine is trained consists of labelled data (both the i/p parameters as well as the reqd o/p.)

- SML Algos. can be broadly divided into 2 types of algos. :-
  - → Classification :- involves predicting a class label. eg. Image classificat", Spam filtering etc
  - → Regression :- involves predicting numerical label. eg. Predicting rainfall based on certain parameters like humidity, temp., windspeed etc
- Egs. - Linear Regression, Logistic Regression, KNN classification, SVM, Decision Trees, Random Forest, Naive Bayes theorem

Unsupervised learning :- Unlike SL algos, where we deal with labelled data for training, the training data will be unlabelled for USML Algos. The clustering of data into a specific grp will be done on the basis of the similarities b/w the vars.

- → Clustering :- involves finding grps. in data.
- → Density Estima" :- involves summarising the distribu" of data
- → Visualiza" :- involves creating plots of data.
- → Projec" :- Involves creating lower dimensional representa" of data

- Egs. - K-means clustering, neural n/ws.
- Egs. - A set of fruits is first fed into the system, & based on its features like color, size, surface type, it categorises them.
  - → finding grps of customers with similar behavior.

Clustering :- Clustering means grouping objs. based on the info found in the data, describing the objs. or their relationship.

- The goal is that objs. in one grp. will be similar to one another & diff. from objs. in another grp.
- It deals with finding a str. in a collec" of unlabelled data.
- The purpose of clustering is to make sense of extract value from large sets of structured & unstructured data.

- Egs. of clustering methods - K-means clustering
  - Fuzzy/C-means clustering
  - Hierarchical clustering.

K-Means Clustering :- aims to identify the best 'k' cluster centers in an iterative manner.
- cluster centers are served as "representative of the objs. associated with the cluster.
- The clusters are assumed to be spherical (drawback).

OPTICS Clustering :- Density based clustering algos.
- It identifies 'dense' cluster of pts., allowing it to learn clusters of arbitrary shape & densities.
- It can also identify outliers (noise) in the data by identifying scattered obj's.

Dimensionality Redn :- As more features are added, the data becomes very sparse, & analysis suffers from the curse of dimensionality.
- Its easier to process smaller data sets.
- Dimensionality redn can be executed in 2 ways :-
  * Feature Selecn :- Selecting from the existing features.
  * Feature Extracn :- Extracting new features by combining existing ones
- - The main technique of feature extracn is PCA (Principal Component Analysis). It guarantees finding the best linear transforman that reduces the no. of dimensions with a min. loss of info..

Reinforcement Learning :- In RL, there's an agent that interacts with a certain env., thus changing its state, & receives reward/penalties for its i/p. Its not told which action is the correct one to achieve its goal.
- Its goal is to find patterns of actions, by ~~trai~~ trying them all & comparing the results, that yield the most reward pts.
- One of the key features of RL is that the agent's actions might not affect the immediate state of the env. but impact the subsequent ones. So, sometimes, the machine doesn't learn whether a certain acn is effective until much later in the episode.
- Egs - Q-learning, temporal-diff. learning, & deep reinforcement learning.
- Egs. Self driving Cars.

Applications of RL :-

Robotics - Google has cut its energy consumpⁿ by about 50% after implem... Deep Mind's Technologies.

Text Mining - Salesforce used RL along with an advanced contextual generaⁿ model to develop a system that's able to produce highly readable summaries of long texts.

Trade Execuⁿ - JPMC announced that it would start using a robot for trading execuⁿ of large orders.

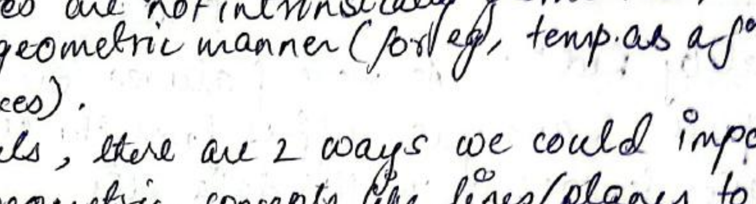
Healthcare - Medicaⁿ dosing, optimizaⁿ of treatment policies for those suffering from chronic, clinical trials, etc.

Learning Models :
- → Geometric Models
- → Probabilistic Models
- → Logical Models

Geometric Models :-

- In geometric models, features could be described as pts. in 2D or a 3D. This is solved by Linear or Non-linear Regression.
- Even when features are not intrinsically geometric, they could be modelled in a geometric manner (for eg, temp. as a fⁿ of time can be modelled in 2 axes).
- In geometric models, there are 2 ways we could impose similarity
  - we could use geometric concepts like lines/planes to segment (classify) the instance space. These are called Linear models.
  - Alternatively, we can use geometric notion of distance to represent similarity. In this case, if 2 pts. are close together, they have similar values for features & thus can be classified as similar. We call such models as Distance-based models.
- It includes
  - Linear Regression
  - Linear Classificaⁿ
  - SVM
  - Nearest Neighbour
  - Clustering.

① Linear Regression :-



→ Minimize sum of squared error.

$$\sum (t-y)^2$$

t: actual value, y: predicted value.

$y = mx + c$

$\Rightarrow y = \omega_0 + \omega_1 x$

- ω_0 : y-intercept
- ω_1 : slope
- ω_0, ω_1 : regression coefficients.
- y : response var.
- x : single predictor var.

Model :- In linear regression, we use a linear f^n of the features $x=(x_1,\ldots,x_D)\in\mathbb{R}^D$ to make prediction's 'y' to the target value 't' $\in\mathbb{R}$.

$$\boxed{y=f(x)=\sum_j w_j x_j + b}$$

- $y \Rightarrow$ prediction
- $w \Rightarrow$ weights associated with each feature.
- $b \Rightarrow$ bias (or intercept).
- 'w' & 'b' together $\Rightarrow$ parameters.

- We hope that our prediction is close to the target ($y \approx t$).

Loss Fun^n :-

- A loss f^n $\mathcal{L}(y,t)$ defines how bad it is if, for some eg. 'x', the algo. predicts 'y', but the target is actually 't'.
- Squared loss f^n : $\boxed{\mathcal{L}(y,t)=\frac{1}{2}(y-t)^2}$ → $(y-t)$ is residual, & we want this small in magnitude.
  → $\frac{1}{2}$ factor is just to make the calc^{n}s convenient.
- Cost f^n: loss f^n averaged over all training egs.

$$J(w,b)=\frac{1}{2N}\sum_{i=1}^{N}\left(y^{(i)}-t^{(i)}\right)^2$$
$$=\frac{1}{2N}\sum_{i=1}^{N}\left(w^T x^{(i)}+b-t^{(i)}\right)^2$$

→ Cost is also called empirical or avg. loss.

Vectorization :-

- The prediction for one data pt. can be computed using a 'for' loop:

```
y=b
for j in range(M):
    y+= w[j]*x[j]
```

- Excessive super/sub scripts are hard to work with, & Python loops are slow, so we vectorize algos. by expressing them in terms of vectors & matrices

$$w=(w_1,\ldots w_D)^T \qquad x=(x_1,\ldots\ldots x_D)^T$$

$$\boxed{y=w^T x+b}$$

- $b \Rightarrow$ bias
- $w^T x \Rightarrow$ dot product of wts. associated with attributes

- This is simpler & executes much faster: $\boxed{y=np.dot(w,x)+b}$.