

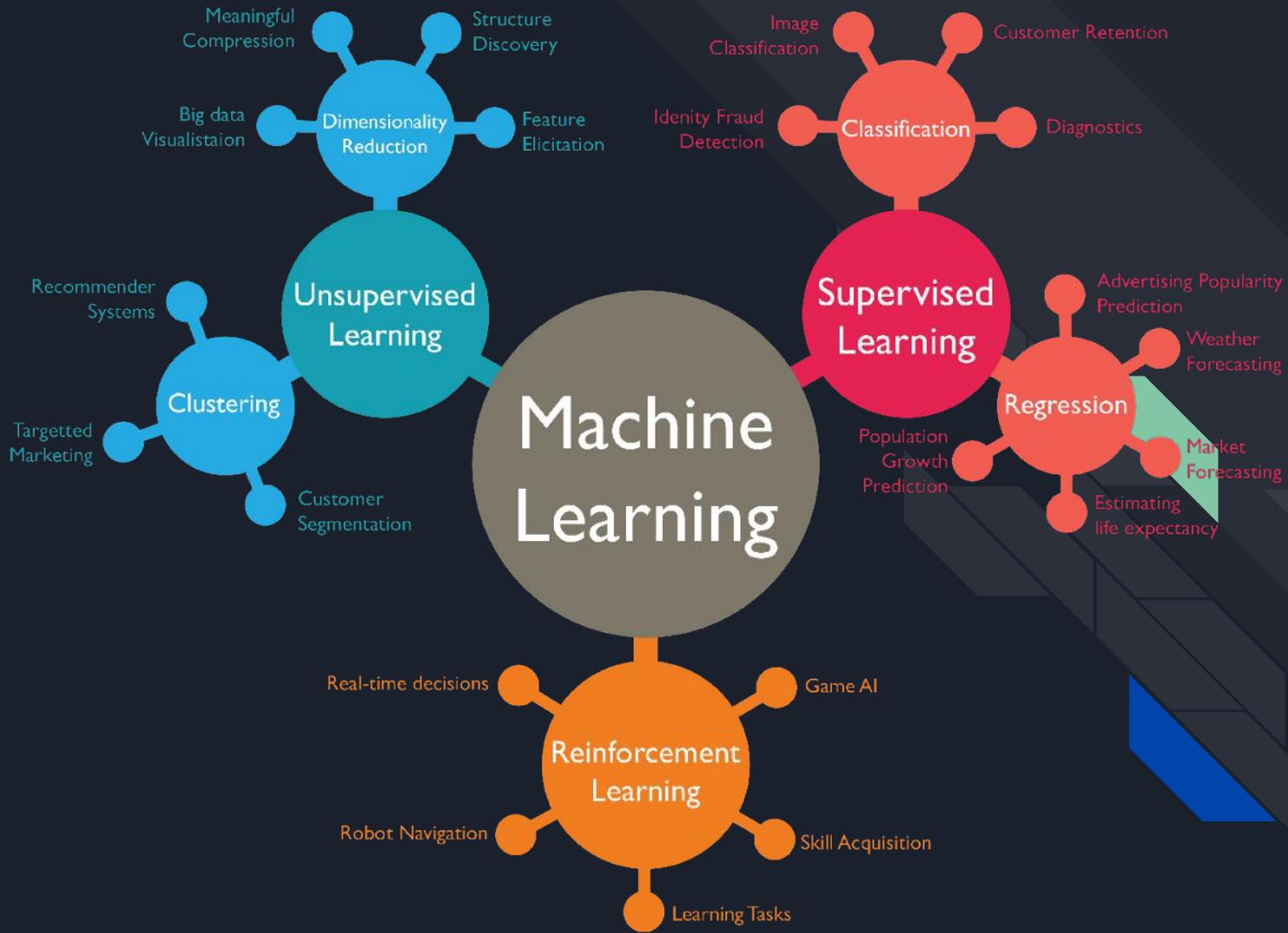


# Machine Learning

## Topics:

- Introduction to Machine Learning
- KNN
- Data Cleaning and Preprocessing

**NOTHING  
WORTH HAVING  
COMES EASY**





# Where you can use ML

- The use case applications for Machine learning are very varied. It is used in all industries in some form or another.
- Supervised machine learning algorithms are widely used in banking sector for fraud detection.
- State of the art supervised learning algorithms over images are widely used in the medical industry for cancer/tumor detection
- Unsupervised algorithms have found a huge variety of applications from things like recommender systems to even medical fields like genomics
- The reinforcement algorithms have mostly found their use in AI development for games and robotic simulations.

# K-Nearest Neighbours

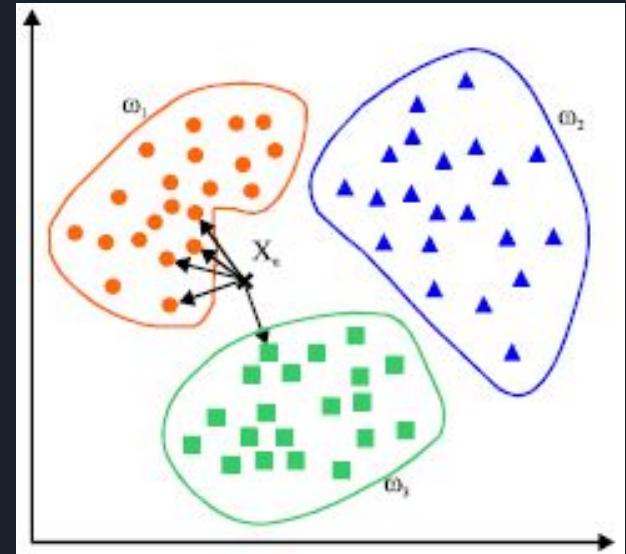
1. It is a supervised machine learning algorithm
2. Generally used for classification
3. It works based on minimum distance from the query instance to the training samples to determine the K-nearest neighbors.

Advantages:

- Simple to implement
- Makes no prior assumption of data

Disadvantages:

- Time complexity is high as distance with all points need to be calculated
- Sensitive to scaling of data (Ye baad me kabhi samajha denge :P)



$$\begin{aligned} d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} \\ &= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}. \end{aligned}$$



# Face Recognition using KNN and HaarCasacade Classifier

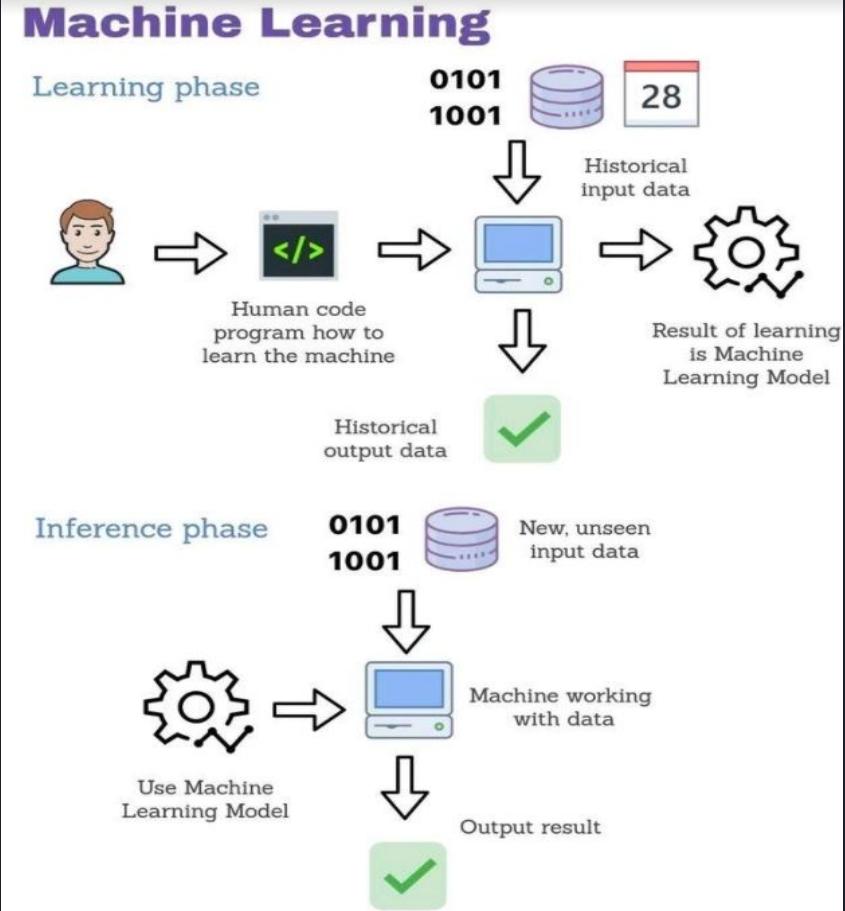
- We can use KNN and Haarcascade for Face Recognition.
- <https://github.com/m1-key/Face-Recognition>
- Haarcascade is a machine learning tool that returns us the coordinates of a rectangular box surrounding our face .
- Now we apply KNN on that and this helps in Face recognition.

- What is data?
- What are data points & features of a dataset?

The diagram illustrates a dataset structure. At the top, 'Features' and 'Labels' are grouped under a bracket above the table. Below this, a green arrow labeled 'Data Point' points to the first row of the table. To the right, a large brace groups the first nine rows as 'Training Data', and the last three rows as 'Test Data'. A red arrow labeled 'NULL data' points to the empty cell in the fourth column of the first row.

Features				Labels
Outlook	Temperature	Humidity	Windy	Play Tennis
Overcast	Hot	High	FALSE	Yes
Overcast	Cool	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Rainy	Cool	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Rainy	Mild	High	FALSE	No
Sunny		High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	
Sunny	Cool	Normal	TRUE	
Sunny	Mild	High	TRUE	

# How are we going to use this data to make machine learn?



Thank you!

