

① ~~Q~~ What is ML?

① It is field of study that gives computers ability to learn without being explicitly programmed.

② ML is training of model from data that generalizes a decision against a performance measure.

② technique →

A tech is a way of solving problem.

1) Supervised

2) Un

3) Reinforcement

① Supervised :- 1) It is similar to human learning in presence of supervisor.

2) Supervisor's roll is to provide correct feedback to learner.

3) • Input to ML model is data & its various attributes are properties.

• along with data the current output is also provided.

4) The aim of Supervised machine learning is to build a model that makes prediction based on evidence. in presence of uncertainty.

5) A supervised learning algorithm takes known set of input data & known set of output data & trains model to generate reasonable response prediction for response to new data.



6) algorithms used in supervised learning :-

- 1) Nearest Neighbour classifier
- 2) Naive Bayes
- 3) Decision Tree
- 4) linear Regression
- 5) Support vector Machine
- 6) Neural Networks.

② Unsupervised :-

- 1) This is learning without teacher.
- 2) This is basically human's ability to group similar element.
- 3) Important characteristic of unsupervised learning is to find similarity bet<sup>n</sup> two events/objects.
- 4) Unsupervised learning finds hidden patterns in data.
- 5) Unsupervised learning is where you only have input data & no corresponding output variable.
- 6) The goal for unsupervised learning is to model the underlying structure / distribution in the data in order to learn more about data.
- 7) These are called unsupervised learning because unlike supervised learning there is no correct answers & there is no teacher.



M T W T F S S  
Page No.:  
Date: / /

13) What is feature eng? What are the steps in feature eng?

- ① Feature eng. is the process of using domain knowledge of data to create features that make ML algorithms work.
- ② Feature eng is process of transforming data into features that better represent the underlying problem, resulting in improved ML performance.
- ③ Before using feature to train machine learning model, it is necessary to clean, transform & select right set of features.

\*Steps:-

1] Preparing data:-

① preparing data takes into account capturing data, storing data, cleaning data, organizing data & so on.

② cleaning data refers to process of transforming data into a format that can be easily interpreted by database.

2] Data cleaning:- Handling missing data

① If a feature having most of missing value, then it can be removed from dataset.

② A missing value can be replaced by mean/median value of that feature/zero value.

3] Handling Numerical features:-

① Min-Max Normalization:-

This process brings the feature value in range. A numerical features holds numerical value.

Refer  
Q.23  
OneTech  
missing



of 0 to 1.

$$X_{\text{new}} = (x_i - \min(x)) / (\max(x) - \min(x))$$

② Z-score Normalisation:-

$$Z = (x - \text{mean}(x)) / \text{stdDev}(x)$$

③ Binning :- A process of converting numerical value into categories

ex. 5, 12, 7, 9, 25, 32, 21, 35

Age	cate
0-10	0
11-20	1
21-30	2
> 31	3

4] Categorical encoding:-

It converts textual categorical values in numbers.

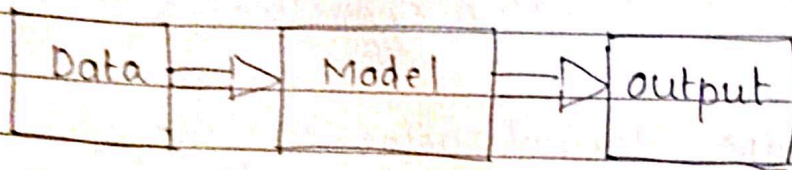
Two Tech. → 1] Label encoding  
2] one hot encoding

1] Label → In this tech., each label is assigned a unique integer based on alphabet ordering.

2] one-hot → It simply creates additional features based on number of ~~unique~~ unique values in categorical feature.  
• every unique value will be added as feature.



## ① ML architecture :-



### ① Data :-

- 1) Data forms the main source of learning in Machine learning.
- 2) Data is representation of human experience in machine learning system.
- 3) Data can be any format - structured  
 - semi-structured  
 - unstructured
- 4) Data can be received at any frequency can be static or dynamic.
- 5) Data can be of any size
- 6) Data can have any dimensions.

### ② Model -

- 1) Model is the representation of real life object. It mimics behaviour of the object it represent.
- 2) A simplified description, especially a mathematical one, of a system or process, to assist calculation and prediction.
- 3) There are 3 categories of model -
  - ① logical
  - ② geometric
  - ③ probabilistic



23] What is data cleaning? explain tech. used for data cleaning?

Refer que. NO. 13 from step

Techniques:-

① Removing unwanted observation

- unwanted obs. like duplicate or irrelevant
- Duplicate obs. most frequently arise during data collection such as:-

- 1) combine dataset from multiple places
- 2) scrape data
- 3) Receive data from clients

② Handling missing data - Refer Q. No. 13

③ categorical encoding → Refer Q. No. 13

① ML Def<sup>n</sup>, arch.

② Techniques

③ Features

④ examples on Recall, precision, accuracy

Q.5A] Accuracy →

Predicted class	Actual class	
	TP	FP
	FN	TN

• Confusion matrix :-

- ① A confusion matrix is a table that is often used to describe the performance of classification model on set of test data for which the true values are known.
- ② It allows the visualization of performance of an algorithm.
- ③ It allows easy identification of confusion bet<sup>n</sup>. classes.
- ④ A confusion matrix is summary of prediction results on classification problem.
- ⑤ The number of correct & incorrect prediction are summarized with count values & broken down by each class. This is the key to confusion matrix.
- ⑥ The confusion matrix shows the ways in which your classification model is confused when it makes predictions.