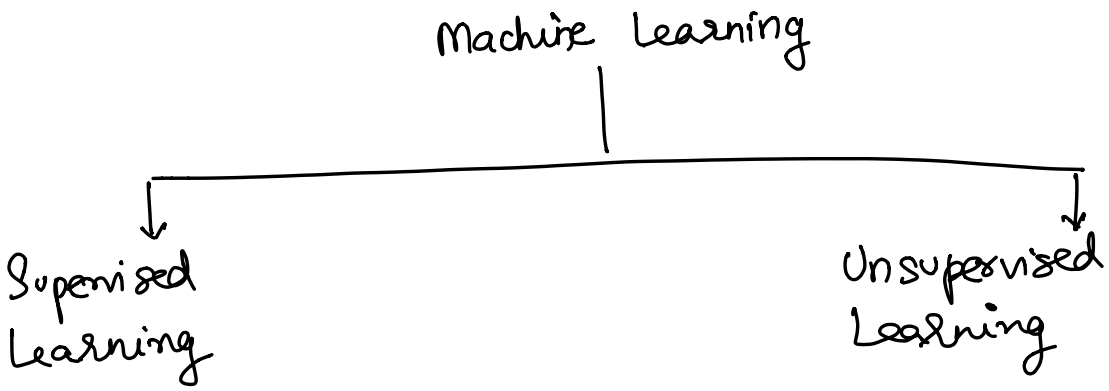
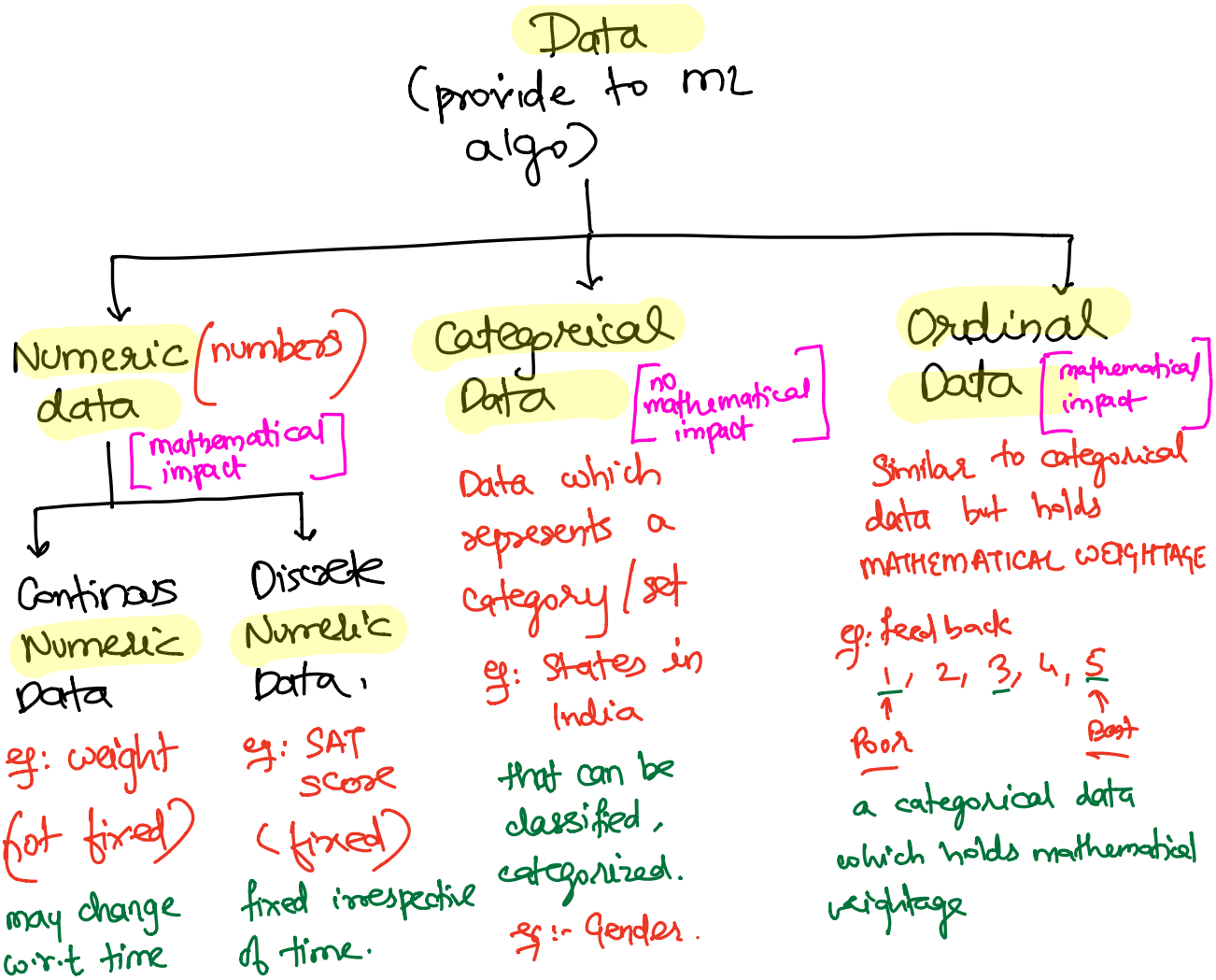
Machine Learning. → Static Model.





WORKFLOW OF A TYPICAL ML ENGINEER

ⓐ DATA PREPROCESSING

ⓑ IDENTIFY WHICH ALGO TO BE USED

ⓒ CREATE MODEL

ⓓ TEST MODEL

ⓔ VALIDATE MODEL

ⓕ FINE TUNE MODEL (IF REQUIRED)

ⓖ DEPLOY MODEL → Check Application Compatibility with model → NO

Data acquisition → Data validation → Data Preprocessing → EDA → Feature Engineering → Feature Enhancement → if NLP (applicable) → Identify Algo → CREATE model → Test model → VALIDATE MODEL → FINE TUNE → Revalidate model → DEPLOY

Data Preparation Phase

This phase is used to make your data compatible to your machine learning algorithm.

① YOUR DATA MUST BE STRICTLY NUMERIC.

SCI-KIT Learn --- package to implement ML algo.

data must be numeric and in the form of numpy array

Handling Categorical data (Handle location col)

eid	esal	location
1	1000	Mumbai
2	2000	Mumbai
3	3000	Pune.

Encode col^n values with numeric data

① location.unique() and store it in the form of list

['Mumbai' | 'Pune']

② Sort the list in ascending order

['Mumbai' | 'Pune']

0 1

③ Replace values of column by the index value of the list

✓ ✓ ✓ ✓ Mathematical weightage (should not happen)

eid	esal	location
1	1000	0
2	2000	0
3	3000	1

The above logic/process is called as Encoding values

Problem → Earlier location didn't contribute any mathematical weightage. Therefore we need to nullify the mathematical weightage

Hot Encoding algo → Remove mathematical weightage using binary transpose

eid	esal	location
1	1000	0
2	2000	0
3	3000	1

⇒

esal	eid	Mumbai	Pune
1000	1	1	0
2000	2	1	0
3000	3	0	1

dummy variables

Dummy variables is meant to nullify the mathematical weightage.