

Feature Engineering

Feature Engineering is the process of using domain knowledge to extract features from raw data.

These features can be used to improve the performance of machine learning algorithm.

Feature Engineering

Feature Transformation

Feature Construction

Feature Selection

Feature Extraction



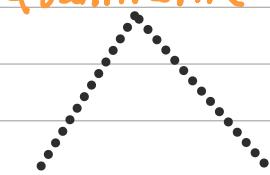
DATA

[+ - × /]

Numerical



Quantitative



Categorical Data

fixed no of categories

Qualitative

Nominal

Ordinal



No Rank of Categories

Rank of Categories

e.g.: Gender

Blood Group

Pincode

e.g.: Customer feedback

Good, Bad, Excellent

e.g.: Bank Account
children

e.g.: Weight
Height, Temp

Handling Missing Data

Missing values occurs in dataset when some of the information is not stored for a variable.

There are 3 mechanism

- missing Completely at Random MCAR
- missing at Random
- missing data not at Random

missing Completely at Random MCAR :

MCAR is a type of missing data mechanism in which the probability of value being missing is unrelated to the observed data and missing data

- no specific or systematic reason for the data to be missing
- missing data are randomly distributed throughout the dataset

- missing at Random :

MAR is a type of missing data mechanism in which the probability of the value being missing depends only on the observed data but not on the missing data itself.

- the missing values are systematically related to observed data, but not to the missing data

missing data not at Random

It is a type of missing data mechanism where the probability of missing values depends on the values of the missing data itself.

- the missingness is not random and is dependent on unobserved or unmeasured factors that are associated with the missing values.