

25/12/20

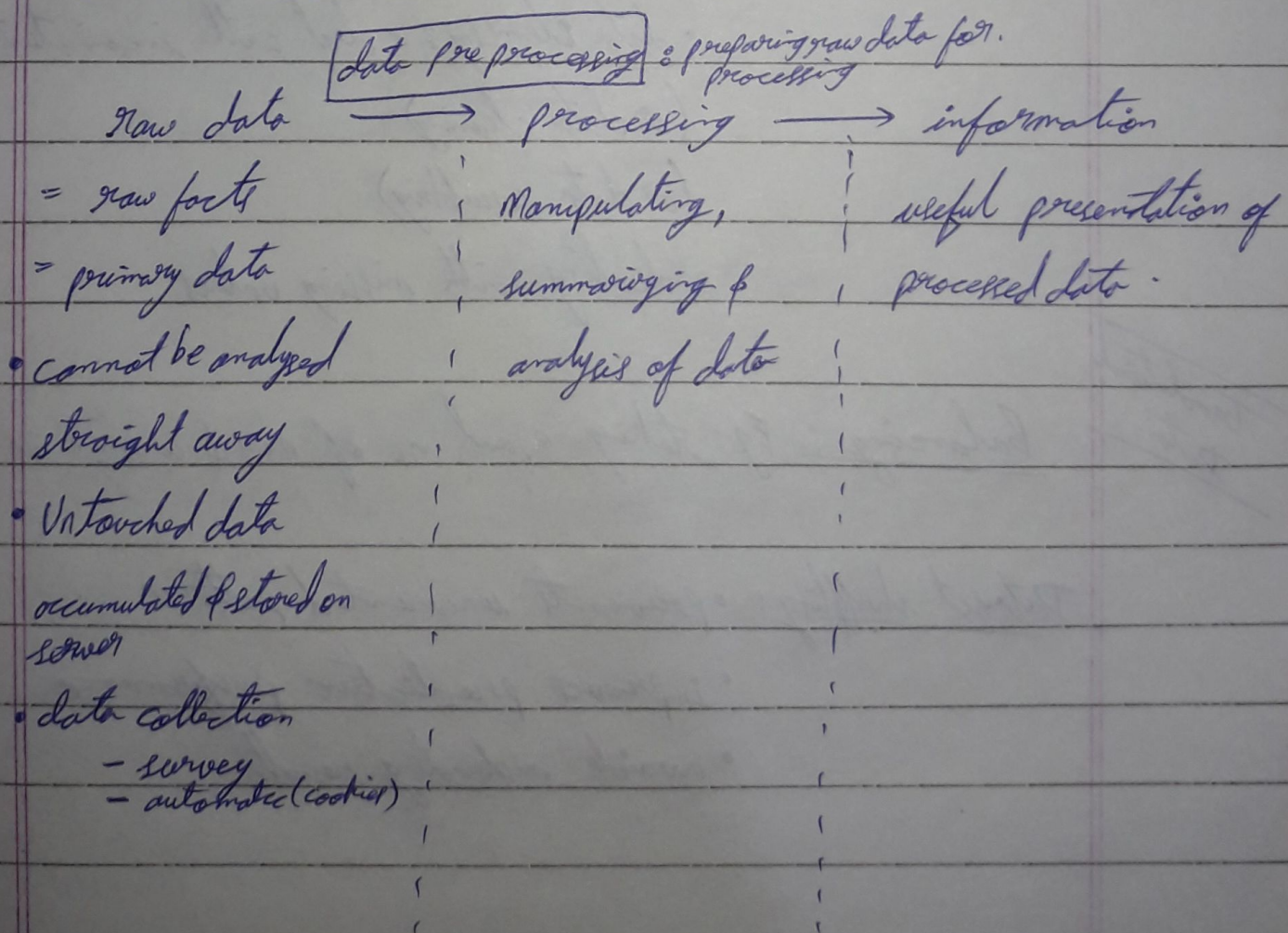
Benefits of Each Discipline

Traditional Methods → Traditional data

Machine Learning → Big Data

Popular Data Science Techniques

1) Techniques with Traditional Data



Traditional Data

data pre processing: • preparing raw data for processing.
• incorrect data must be marked invalid or corrected.

Methods → • class labeling: labeling data to correct type
numerical categorical
can be manipulated cannot be manipulated

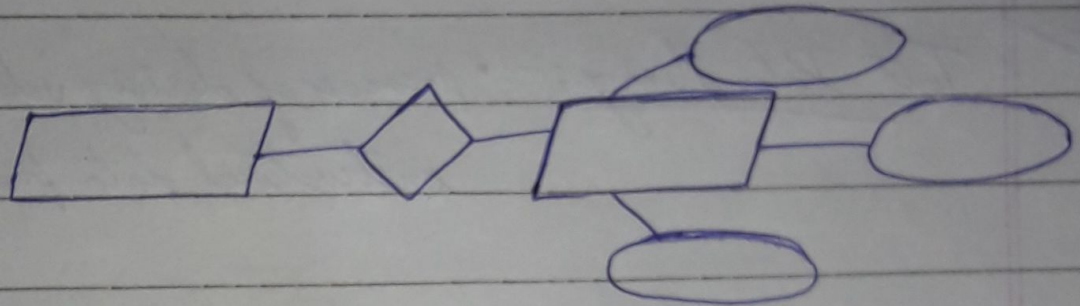
- data cleaning: • deal with inconsistent data
(or data cleaning)
(or data scrubbing)
- dealing with missing values

Traditional Data

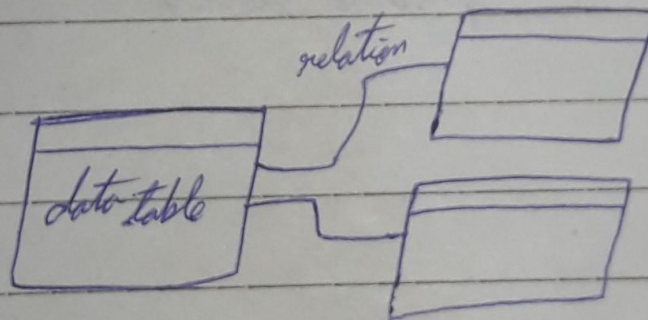
Balancing: Eg: taking equal no of men & women in shopping

Dataset shuffling: • prevents unwanted patterns
• improves predictive performance
• avoid misleading results.

Traditional Data



Entity-Relationship Diagram (ER Diagram)



Relational Schema

Big Data

Pre processing :-

- class labeling

- number, text,

- digital image, video, audio

- data cleansing

- missing values