05 August 2024   10:00

DATA VS INFORMATION

Data

- Raw facts
- Based on records, observations etc
- Unorganised

Information

Processed data is information

TYPES OF DATA

Unstructured: Raw, unorganised

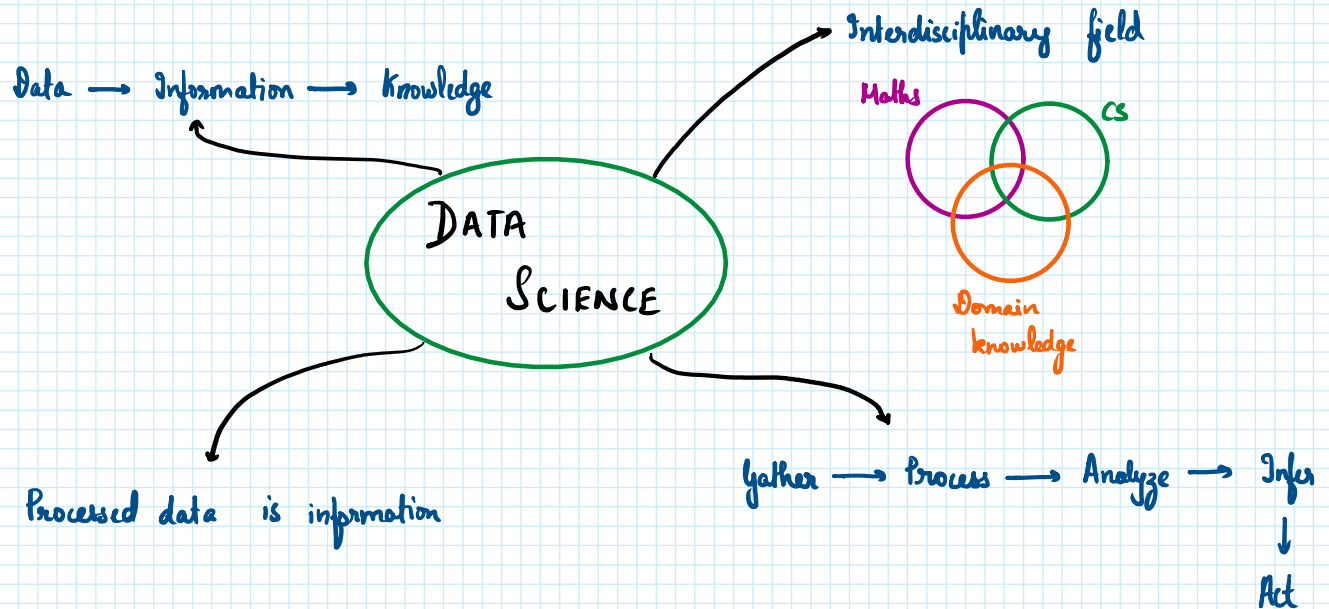Semi-structured: Not in database but has some organisational properties

Structured: Fixed fields, organised

AIM OF DATA SCIENCE

Explore: Identifying patterns in raw data

Infer: Quantify what is learnt from data

Predict: Estimate future trends

Data → Information → Knowledge

Interdisciplinary field

Maths   CS   Domain knowledge

DATA SCIENCE

Processed data is information

Gather → Process → Analyze → Infer ↓ Act

STATISTICAL ANALYSIS

- Make inferences about population from a smaller sample

Population: Entire collection of objects/outcomes undergoing analysis

Sample: Subset of population that represents it for analysis

6σ: standered

Characteristics of samples

- Representative of <u>entire</u> population
- Appropriately sized
- Unbiased
- Economical
- Goal-oriented

Types of population

Tangible

Physical objects/quantities

Eg: Books, people

Conceptual

Quantities that are not discrete physical quantities

Eg: Temperature measurement

TARGET POPULATION
Entire group of elements that comes under analysis

STUDY POPULATION
The population that researchers can apply their conclusion to

SAMPLE