QQ: 17937910

Raw data is useless — analyze and extract — patterns.
(recorded facts)

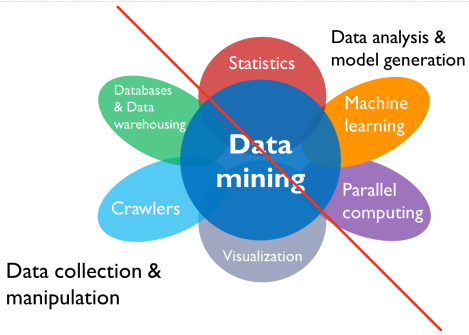
large amount of data ⟶ interesting patterns
(implicit, previously unknow, potentially useful)

1. extract knowledge.
2. large bodies of data

Broad data mining

- Broad data mining include
  - Data collection
  - Data integration
  - Data preprocessing
  - Data visualization — explorative
  - Model generation
  - Model evaluation
  - Model optimization
  - Model deployment
  - Incremental improvement 循环

Related domains



predictive DM

Descriptive DM

Application of DM.

Search, Computer vision, Track analysis of sharing bikes, recommendation Sys.

Machine translation ⟶ NLP & G.

Sports. Economy, Drug development & public health.

Learning objectives of this course

- **Learning goal**: be able to apply the techniques and tools to solve classic data mining problems.
  - **basic concepts**
  - **basic methods**
  - **basic procedures**
  - **common tools**
- **Focus**: practical data mining techniques and tools.

Textbooks

- Data mining: practical machine learning tools and techniques (4th ed., 2016)
  Ian H. Witten, Eibe Frank, Mark A. Hall

Course structure

1. Introduction
2. Programming basics: Python, Numpy & Pandas
3. Data preprocess & analysis
4. Machine learning basics
5. Numeric prediction
6. Category prediction (classification)
7. Cluster analysis
8. Correlation analysis

How to achieve the goal?

- Final exam: Course project
  - Written report
  - Recorded video
  - Demonstration (TBD)
  - Final grade: 20% from your classmates.
- Final grade:
  - in-class performance, assignments, tutorials: 30%
  - final project: 70% (20% from your classmates)

kaggle