นายอภิวัฒน์ เหลี่ยมสิงขร 623020545-4

สิ่งที่รู้ในวันนี้

Why Data Mining?

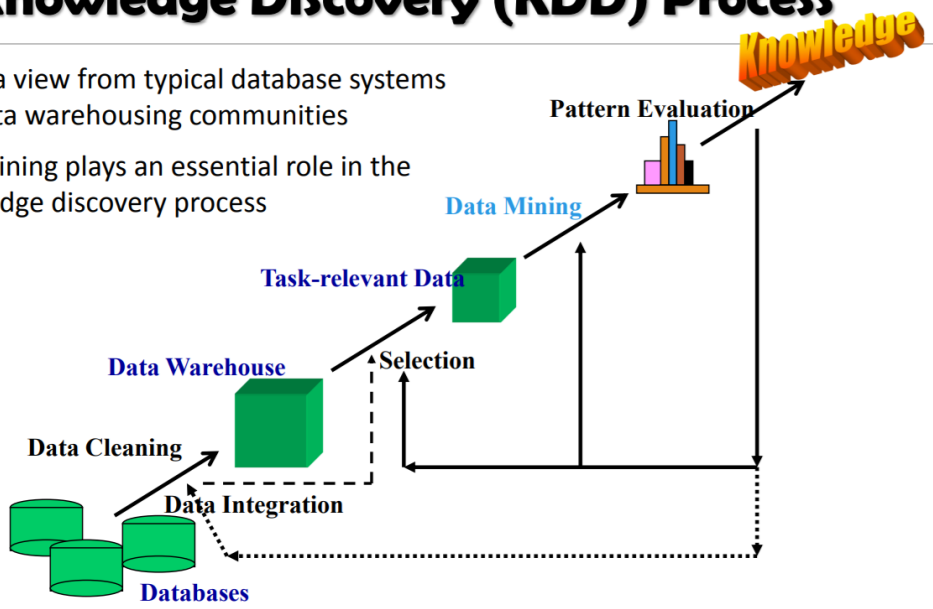
- The Explosive Growth of Data: from terabytes to petabytes
  - Data collection and data availability
    - Automated data collection tools, database systems, Web, computerized society
  - Major sources of abundant data
    - Business: Web, e-commerce, transactions, stocks, …
    - Science: Remote sensing, bioinformatics, scientific simulation, …
    - Society and everyone: news, digital cameras, YouTube
- We are drowning in data, but starving for knowledge!
- "Necessity is the mother of invention"—Data mining—Automated analysis of massive data sets

Data and Information Systems (DAIS:) Course Structures at CS/UIUC

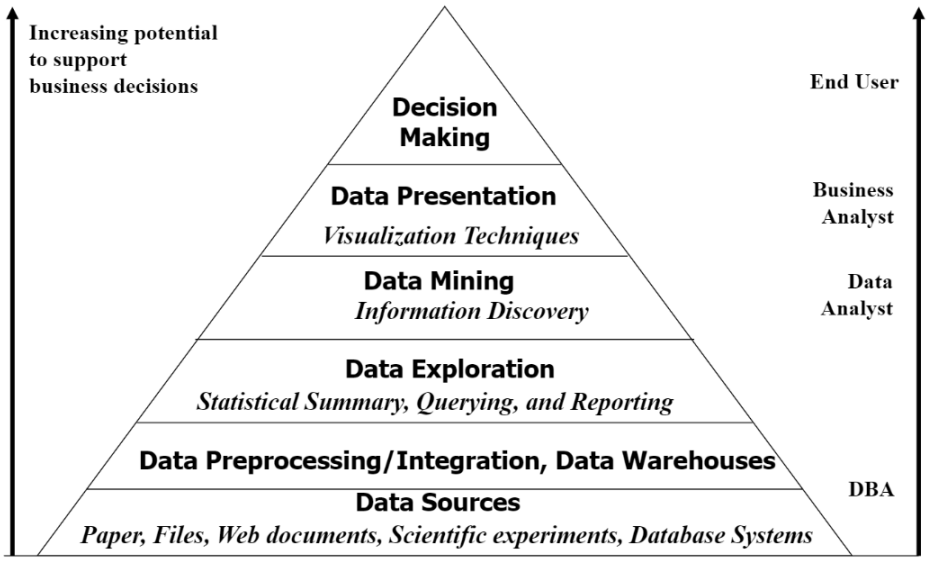
- Coverage: Database, data mining, text information systems, Web and bioinformatics
- Data mining
  - Intro. to data warehousing and mining (CS412)
  - Data mining: Principles and algorithms (CS512)
- Database Systems:
  - Intro. to database systems (CS411)
  - Advanced database systems (CS511)
- Text information systems
  - Text information system (CS410)
  - Advanced text information systems (CS510)

Knowledge Discovery (KDD) Process

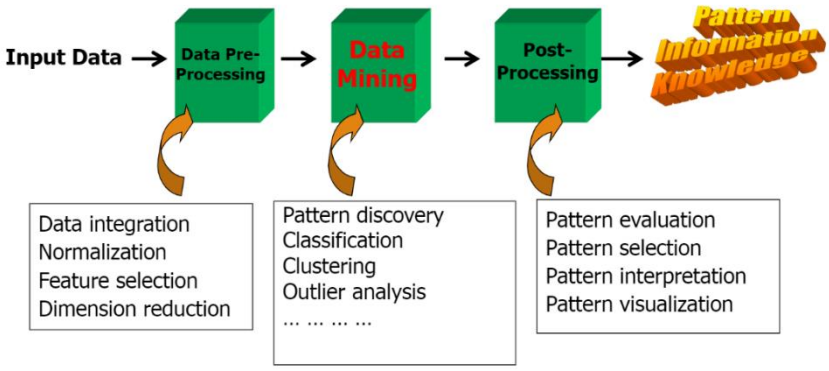
- This is a view from typical database systems and data warehousing communities
- Data mining plays an essential role in the knowledge discovery process



Data Mining in Business Intelligence

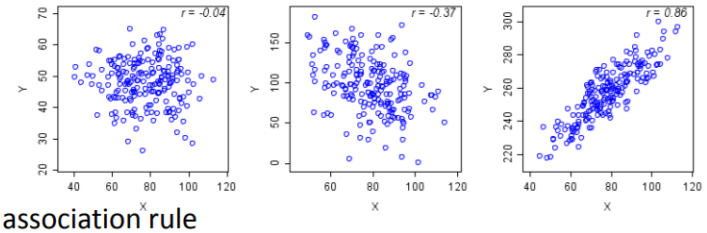


KDD Process: A View from ML and Statistics



Data Mining Functions: (2) Pattern Discovery

- Frequent patterns (or frequent itemsets)
  - What items are frequently purchased together in your Walmart?
- Association and Correlation Analysis



- A typical association rule
  - Diaper → Beer [0.5%, 75%] (support, confidence)
  - Are strongly associated items also strongly correlated?
- How to mine such patterns and rules efficiently in large datasets?
- How to use such patterns for classification, clustering, and other applications?