

Unit 1

Why Data Mining? What is Data Mining

Why Data mining?

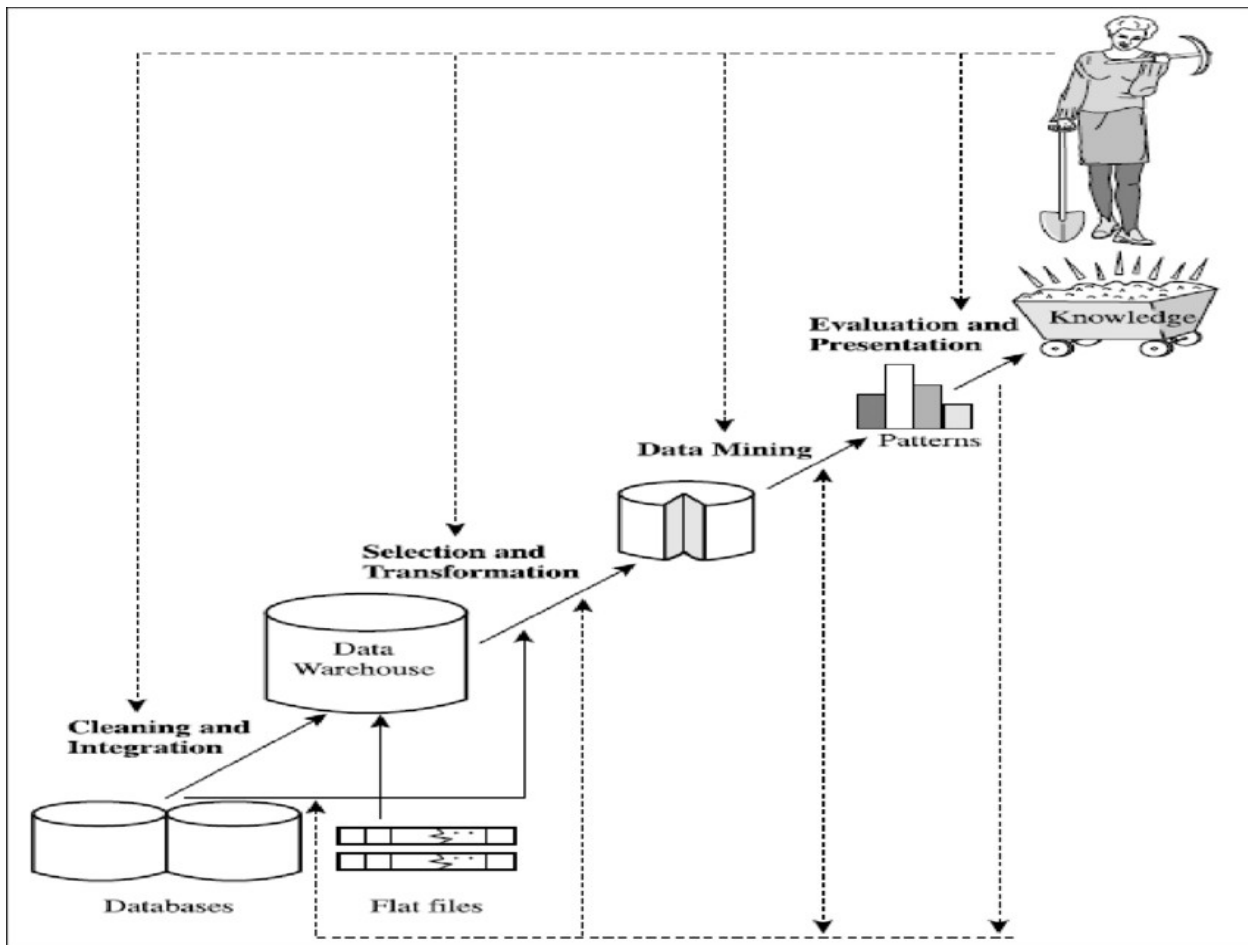
- We live in a world where vast amounts of data are collected daily. Analyzing such data is an important need.
- The explosively growing, widely available, and gigantic body of data makes our time truly the data age.
- Powerful and versatile tools are badly needed to automatically uncover valuable information from the tremendous amounts of data and to transform such data into organized knowledge.
- This necessity has led to the birth of data mining.

What is Data mining?

- Data mining is the process of discovering interesting patterns and knowledge from large amounts of data.
- The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically.

The knowledge discovery process is

1. Data cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations)
5. Data mining (an essential process where intelligent methods are applied to extract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on interestingness measures)
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present mined knowledge to users)



Data mining as a step in the process of knowledge discovery

- Steps 1 through 4 are different forms of data preprocessing, where data are prepared for mining.
- The data mining step may interact with the user or a knowledge base.
- The interesting patterns are presented to the user and may be stored as new knowledge in the knowledge base.
- The preceding view shows data mining as one step in the knowledge discovery process, an essential one because it uncovers hidden patterns for evaluation.
- However, in industry, in media, and in the research milieu, the term data mining is often used to refer to the entire knowledge discovery process.