

The data mining process



Data Base and Data Mining Group of Politecnico di Torino

Elena Baralis
Politecnico di Torino



Data mining

→ is the process of extracting information

■ Non trivial extraction of

- implicit
- previously unknown
- potentially useful

→ information are extracted in a specific way, the process is complex

information from available data

✓ Extraction is automatic

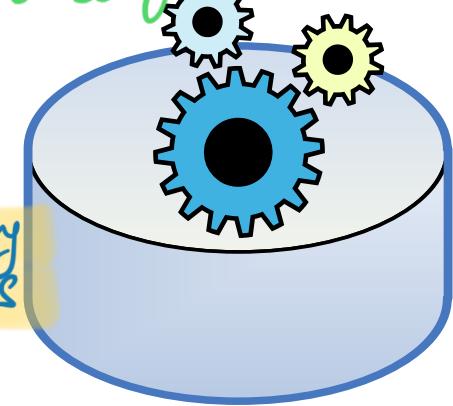
→ done by using proper algorithms

- performed by appropriate algorithms

■ Extracted information is represented by means of abstract models

- denoted as pattern

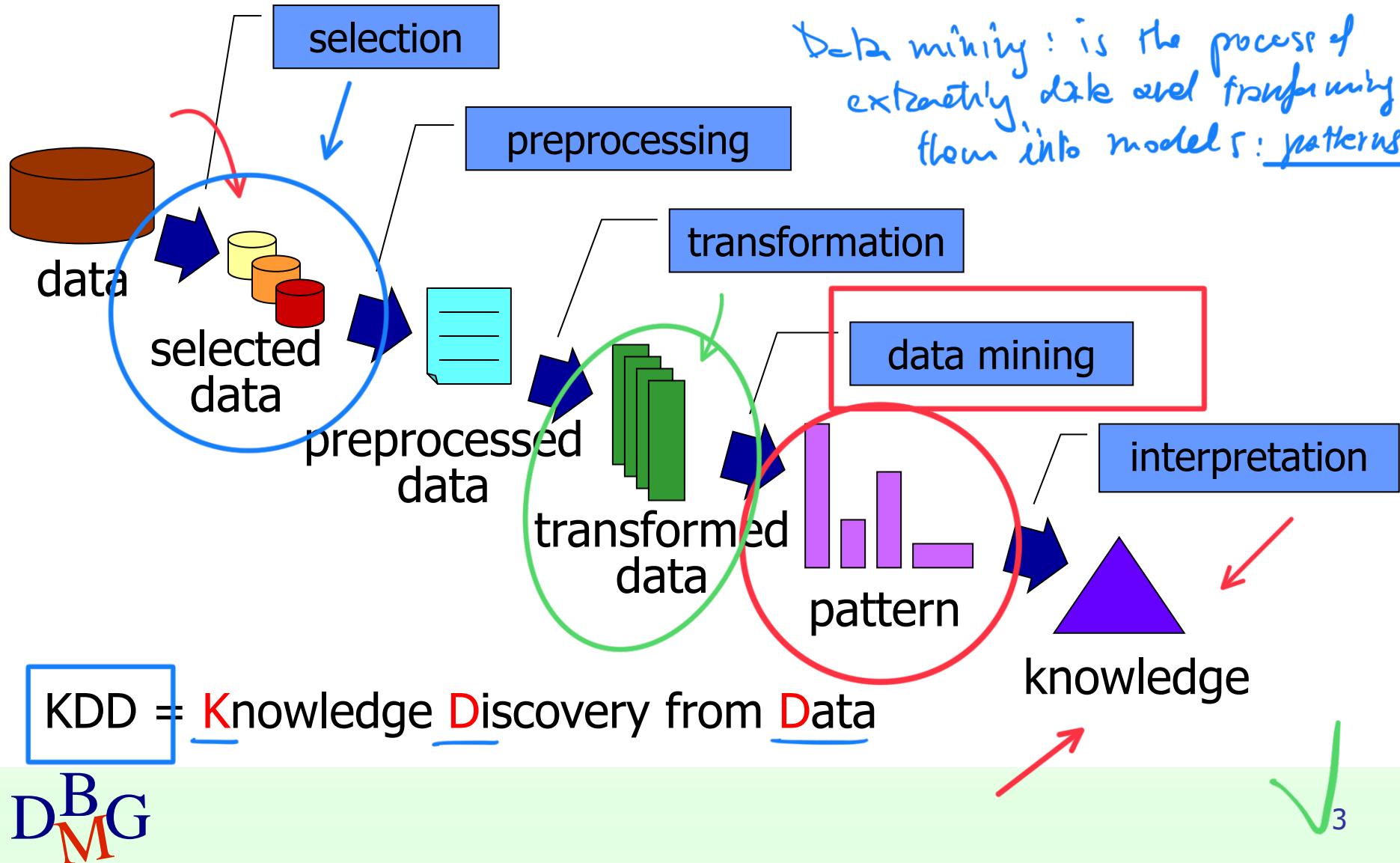
- • aggregation patterns
- clusters
- etc



✓



Knowledge Discovery Process





Analysis techniques

1) Descriptive methods

- Extract interpretable models describing data
- Example: client segmentation

2 techniques:

providing details

2) Predictive methods

- Exploit some known variables to predict unknown or future values of (other) variables
- Example: “spam” email detection





Analysis techniques

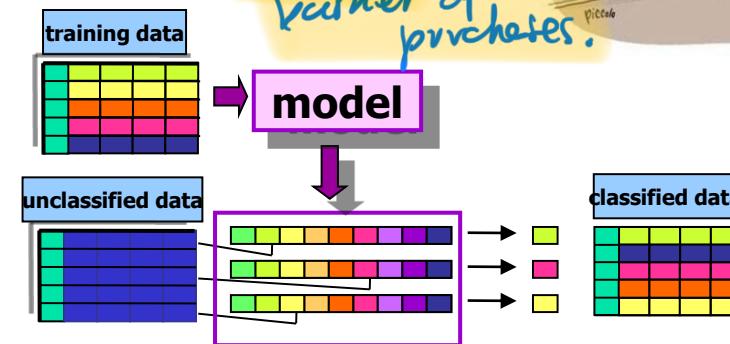
Association rule extraction

→ used to analyze correlations between items
- Descriptive methods
market basket of purchases.



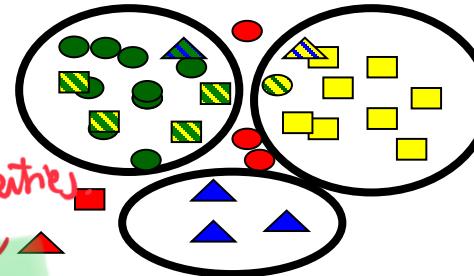
Classification

Analyze a data collection and build a classifier:
an additional CLASS label.
Email: spam or not Spam



Clustering

collection of items
partitioning data into
groups having similar properties
grouped by SIMILARITY



D^B_MG