

Data Mining

Associate Professor Dr. Raed Ibraheem Hamed

University of Human Development,
College of Science and Technology
Department of Computer Science

2016 – 2017

Department of CS- DM - UHD



Introduction



- Multi-Dimensional View of Data Mining
- Data Mining: Classification Schemes
- Data Mining: On what kind of data?
- The Primary Tasks of Data Mining
- Basic Data Mining Tasks
- Applications of Data Mining
- Requirements for a Data Mining System

Multi-Dimensional View of Data Mining

■ Data to be mined

- Relational, data warehouse, transactional, stream, object-oriented/relational, active, spatial, time-series, text, multi-media, heterogeneous, legacy, WWW

■ Knowledge to be mined

- Characterization, discrimination, association, classification, clustering, trend/deviation, outlier analysis, etc.
- Multiple/integrated functions and mining at multiple levels

■ Techniques utilized

- Database-oriented, data warehouse (OLAP), machine learning, statistics, visualization, etc.

■ Applications adapted

- Retail, telecommunication, banking, fraud analysis, bio-data mining, stock market analysis, text mining, Web mining, etc.

Data Mining: Classification Schemes

Different views lead to different classifications:

1. **Data** view: Kinds of data to be mined
2. **Knowledge** view: Kinds of knowledge to be discovered
3. **Method** view: Kinds of techniques utilized
4. **Application** view: Kinds of applications adapted

Data Mining: On What Kinds of Data?

- Database-oriented data sets and applications
 - Relational database, data warehouse, transactional database
- Advanced data sets and advanced applications
 - Data streams and sensor data
 - Time-series data, temporal data, sequence data (incl. bio-sequences)
 - Structure data, graphs, social networks and multi-linked data
 - Object-relational databases
 - Heterogeneous databases and legacy databases
 - Spatial data and spatiotemporal data
 - Multimedia database
 - Text databases
 - The World-Wide Web

The Primary Tasks of Data Mining

The two "high-level" primary goals of data mining, in practice, are *prediction* and *description*.

- 1. Prediction:** involves using some variables or fields in the database to predict unknown or future values of other variables of interest.
- 2. Description:** focuses on finding human-interpretable patterns describing the data.

Predictive Model

- Involves using some variables or fields in the database to predict unknown or future values of other variables of interest.
- Or based on the use of other historical data.
- **Example :-**
 1. Credit card fraud
 2. Breast cancer early warning
 3. Terrorist act

Descriptive Model

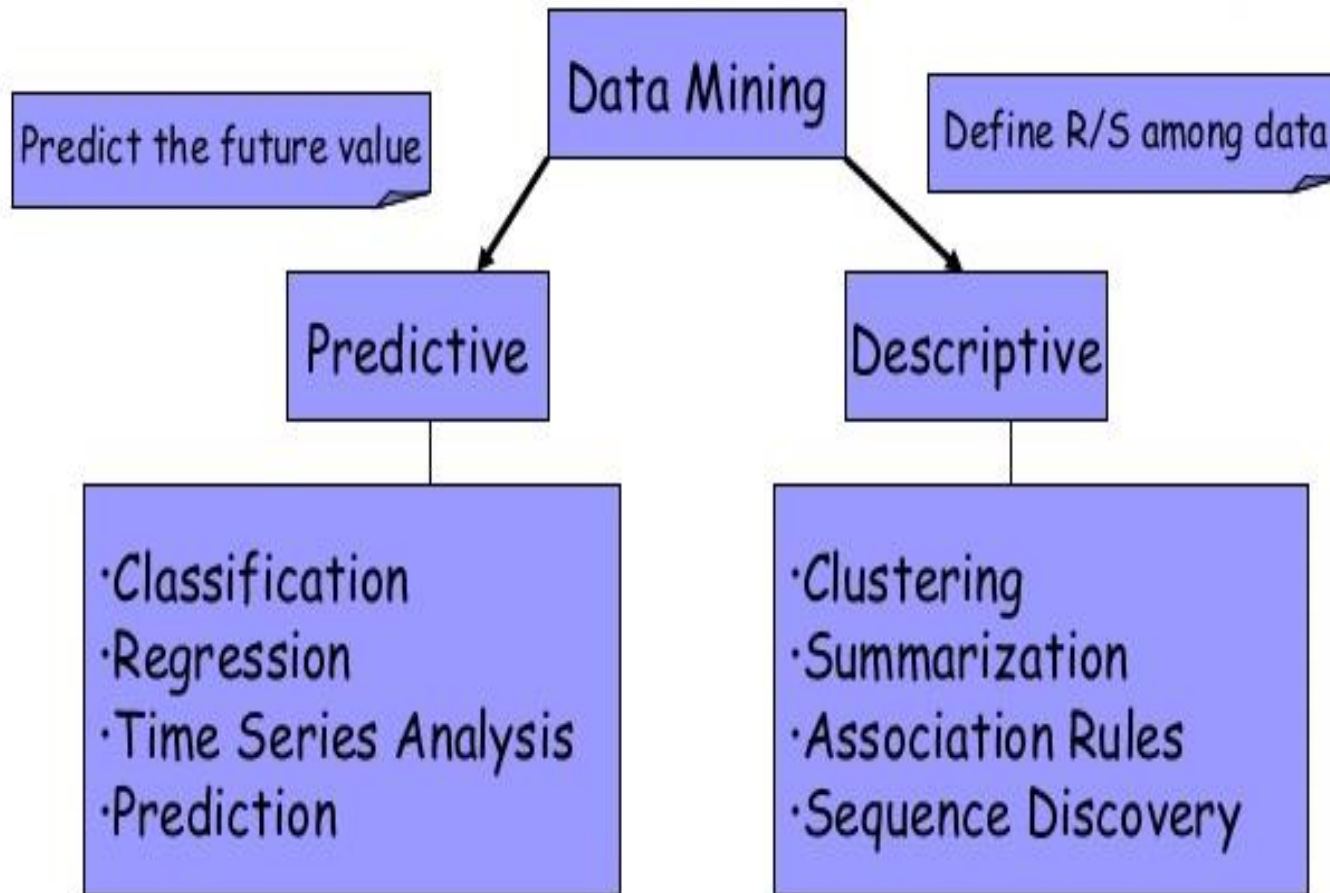
- Focuses on finding human-interpretable patterns describing the data.
- Serves as a way to explore the properties of data examined, not to predict new properties.
- Always required a domain expert.
- **Example:**
 1. Segmenting marketing area
 2. Profiling student performances
 3. Profiling GooglePlay/ AppleApps customer

Basic Data Mining Tasks

The goals of prediction and description are achieved by using the following primary data mining tasks:

- 1) Classification
- 2) Pattern Regression
- 3) Time serious analysis
- 4) Prediction
- 5) Clustering
- 6) Association rules
- 7) Summarization
- 8) Sequence discovery

Basic Data Mining Tasks



Applications of Data Mining

- E-commerce
- Marketing and retail
- Finance
- Telecoms
- Drug design
- Process control
- Space and earth sensing
- Bioinformatics
- Etc.

Requirements for a Data Mining System

- Data mining systems should be:
 - 1) Computationally sound
 - 2) Statistically sound
 - 3) Ergonomically sound



*Thank
you*

