RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-19

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- DATABASE PROCESSING VS. DATA MINING PROCESSING
- QUERY EXAMPLES
- DATA MINING MODELS AND TASKS
- BASIC DATA MINING TASKS
- DATA MINING AND BUSINESS INTELLIGENCE
- MCQ
- REFERENCES

Database Processing vs. Data Mining Processing

❑ **Query**

- Well defined
- SQL

❑ **Data**

- Operational data

❑ **Output**

- Precise
- Subset of database

❑ **Query**

- Defined Poorly
- No precise query language

❑ **Data**

- Not operational data

❑ **Output**

- Fuzzy
- Not a subset of database

Query Examples

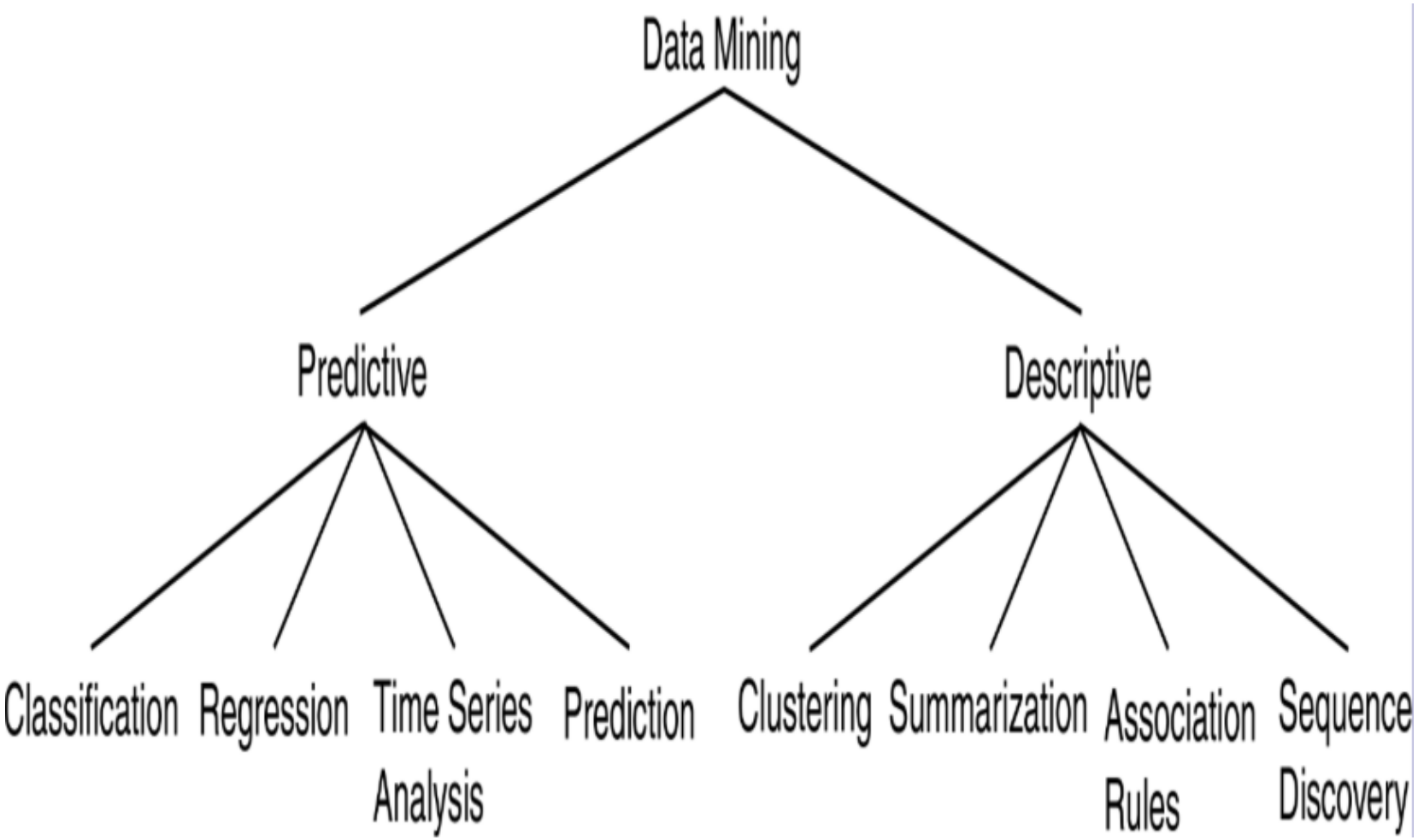
❑Database

- Find all credit applicants with last name as Smith.
    - Identify customers who have purchased more than $10,000 in the last month.
        - Find all customers who have purchased milk

❑Data Mining

- Find all credit applicants who have poor credit risks. (classification)
    - Identify customers with similar buying habits.(Clustering)
    - Find all items which are frequently purchased with
    - milk. (association rules)

Data Mining Models and Tasks



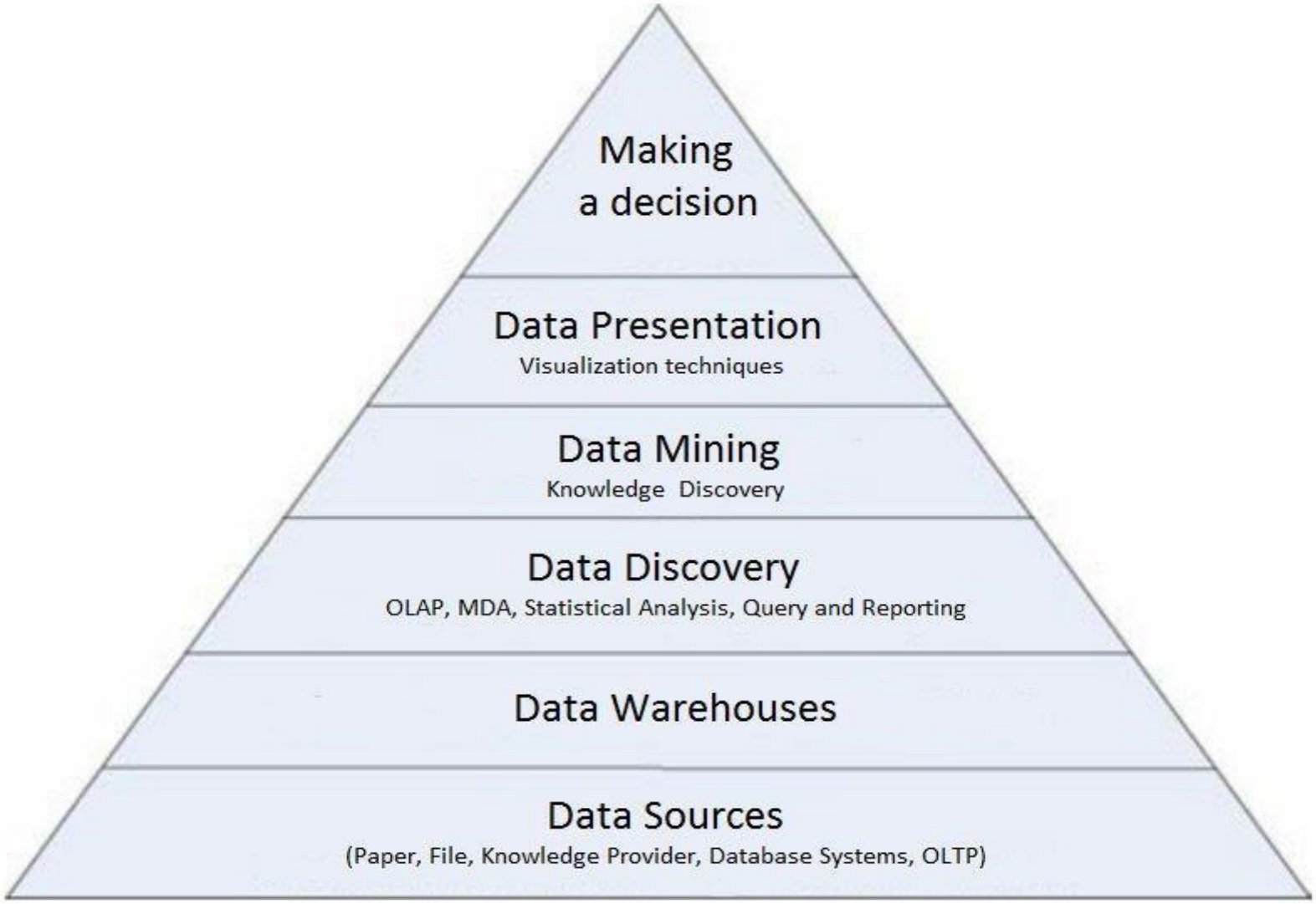
Basic Data Mining Tasks

- **Classification** maps data into predefined groups or classes
  - Supervised learning
  - Pattern recognition
  - Prediction
- **Regression** is used to map a data item to a real valued prediction variable.
- **Clustering** groups similar data together into clusters.
  - Unsupervised learning
  - Segmentation
  - Partitioning

Basic Data Mining Tasks

- **Summarization** maps data into subsets with associated simple descriptions.
  - Characterization
  - Generalization
- **Link Analysis** uncovers relationships among data.
  - Affinity Analysis
  - Association Rules (Finds rule of the form: X=>Y Or " If X then Y")
  - Sequential Analysis determines sequential patterns.
- **(Artificial) Neural Networks**
- **Genetic algorithms**
- **Hypothesis Testing.**

Data Mining and Business Intelligence



Multiple Choice Question

1. The power of self-learning system lies in _____.
a) cost
b) speed
c) accuracy
d) simplicity
e) .

2.. Building the informational database is done with the help of _____.
a) transformation or propagation tools.
b) transformation tools only.
c) propagation tools only.
d) extraction tools.

3. How many components are there in a data warehouse?
a) two
b) three
c) four
d) five

4. Which of the following is not a component of a data warehouse?
a) Metadata
b) Current detail data.
c) Lightly summarized data.
d) Component Key.

5. Highly summarized data is _____.
a) compact and easily accessible.
b) compact and expensive.
c) compact and hardly accessible.
d) compact.

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf DATA MINING BOOK WRITTEN BY Micheline Kamber
- https://www.javatpoint.com/three-tier-data-warehouse-architecture
- M.H. Dunham, " Data Mining: Introductory & Advanced Topics" Pearson Education
- Jiawei Han, Micheline Kamber, " Data Mining Concepts & Techniques" Elsevier
- Sam Anahory, Denniss Murray," data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, " Pearson Education
- Mallach," Data Warehousing System", TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE'97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB'96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD'97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.