RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-20

MR. DHIRENDRA

ASSISTANT PROFESSOR

RAMA UNIVERSITY

OUTLINE

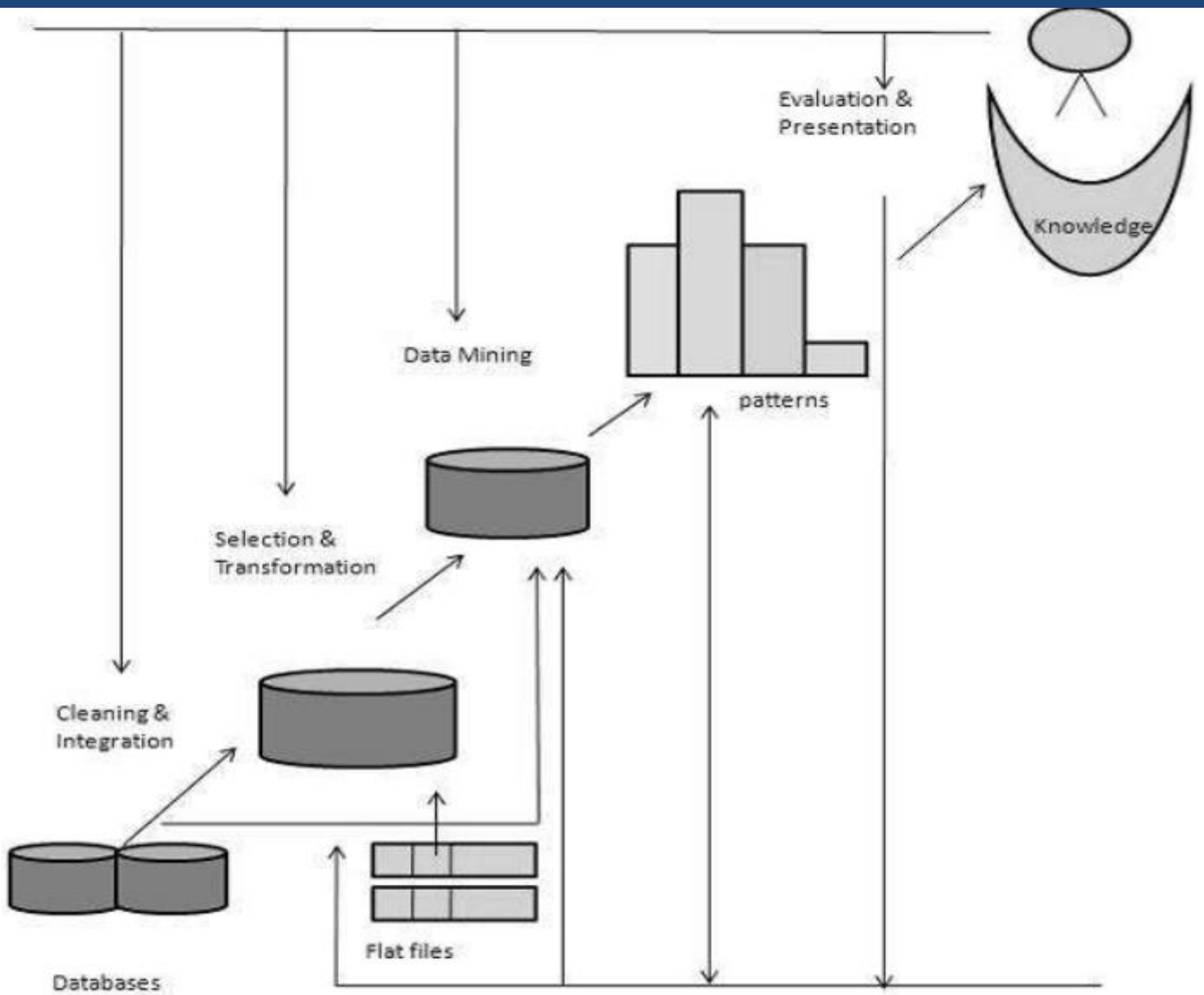
- KNOWLEDGE DISCOVERY
- DATA MINING - SYSTEMS
- DATA MINING SYSTEM CLASSIFICATION
- CLASSIFICATION BASED ON THE DATABASES MINED
- CLASSIFICATION BASED ON THE KIND OF KNOWLEDGE MINED
- CLASSIFICATION BASED ON THE TECHNIQUES UTILIZED
- CLASSIFICATION BASED ON THE APPLICATIONS ADAPTED
- INTEGRATING A DATA MINING SYSTEM WITH A DB/DW SYSTEM
- MCQ
- REFERENCES

Knowledge Discovery

❑ Here is the list of steps involved in the knowledge discovery process −

- **Data Cleaning** In this step, the noise and inconsistent data is removed.
- **Data Integration** In this step, multiple data sources are combined.
- **Data Selection** In this step, data relevant to the analysis task are retrieved from the database.
- **Data Transformation** In this step, data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.
- **Data Mining** In this step, intelligent methods are applied in order to extract data patterns.
- **Pattern Evaluation** In this step, data patterns are evaluated.
- **Knowledge Presentation** In this step, knowledge is represented.

Knowledge Discovery



Data Mining - Systems

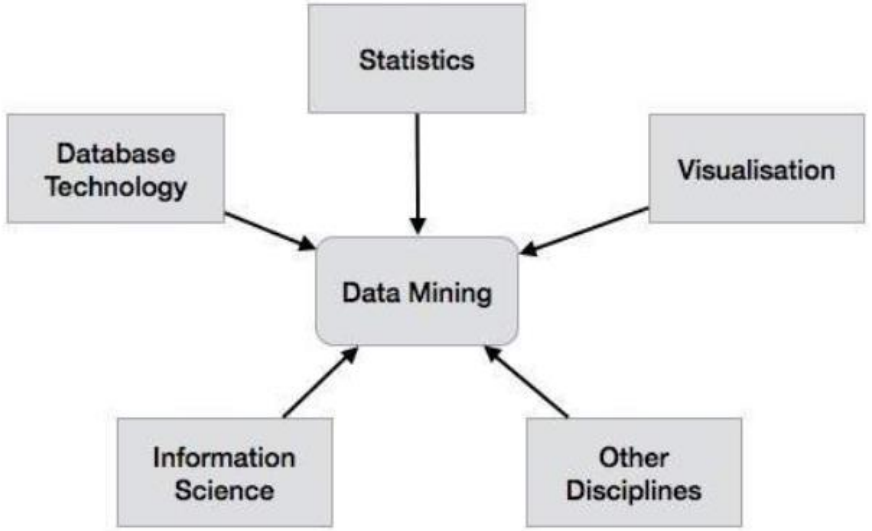
There is a large variety of data mining systems available. Data mining systems may integrate techniques from the following −

- Spatial Data Analysis
- Information Retrieval
- Pattern Recognition
- Image Analysis
- Signal Processing
- Computer Graphics
- Web Technology
- Business
- Bioinformatics

Data Mining System Classification

A data mining system can be classified according to the following criteria −

- Database Technology
- Statistics
- Machine Learning
- Information Science
- Visualization
- Other Disciplines



Classification Based on the Databases Mined

- We can classify a data mining system according to the kind of databases mined. Database system can be classified according to different criteria such as data models, types of data, etc. And the data mining system can be classified accordingly.
- For example, if we classify a database according to the data model, then we may have a relational, transactional, object-relational, or data warehouse mining system.

Classification Based on the kind of Knowledge Mined

We can classify a data mining system according to the kind of knowledge mined. It means the data mining system is classified on the basis of functionalities such as −

- Characterization
- Discrimination
- Association and Correlation Analysis
- Classification
- Prediction
- Outlier Analysis
- Evolution Analysis

Classification Based on the Techniques Utilized

We can classify a data mining system according to the kind of techniques used. We can describe these techniques according to the degree of user interaction involved or the methods of analysis employed.

Classification Based on the Applications Adapted

We can classify a data mining system according to the applications adapted. These applications are as follows −
Classification Based on the Techniques Utilized

- Finance
- Telecommunications
- DNA
- Stock Markets
- E-mail

Integrating a Data Mining System with a DB/DW System

- No Coupling
- Loose Coupling
- Semi−tight Coupling
- Tight coupling

Multiple Choice Question

1. "_____ contains information that gives users an easy-to-understand perspective of theinformation stored in the data warehouse."
a) Business metadata.
b) Technical metadata.
c) Operational metadata.
d) Financial metadata.

2.. _____ helps to integrate, maintain and view the contents of the data warehousing system.
a) Business directory.
b) Information directory.
c) Data dictionary.
d) Database

3. Discovery of cross-sales opportunities is called _____.
a) segmentation
b) visualization
c) correction
d) association

4. Data marts that incorporate data mining tools to extract sets of data are called _____.
a) independent data mart.
b) dependent data marts.
c) intra-entry data mart.
d) inter-entry data mart.

5. _____ can generate programs itself, enabling it to carry out new tasks.
a) Automated system.
b) Decision making system.
c) Self-learning system.
d) Productivity system.

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.