

Unit 1:

Introduction to Data Mining; Functionalities, Kinds of Data, Classification of Data Mining, Data Mining Primitives, Data Warehousing and Schemas

Unit-2: Data Pre-Processing

Unit-3: Mining Frequent Itemsets and Association Mining

Unit-4:

Classification-1 : Decision Tree Induction, Bays Classification, Multi-Layer Feed Forward Neural Network

Unit-5:

Classification-2: Support Vector Machine and Regression

Unit-6:

Cluster Analysis

Text Book

“Data Mining Concepts and Techniques”- Third Edition,
Jiawei Han and Kamber

Other References

“Data Mining Techniques” – Arun K Pujari, University
Press

1. What is Data Mining?

“Data mining refers to extracting or mining knowledge from large amount of data”

It focuses on extraction of useful information

It transform the data into interpretable structure for further use

Analyze the data with data mining functionalities for data/pattern analysis.

Introduction : Data Mining



2. Motivation of Data Mining

Access and extract the required data from either peta or terabyte amounts of data and this feature is motivated with following studies

1. Banking, insurance, financial transactions - electronic banking, ATMs, credit cards, stock market data
2. Supermarket check-out scanner data, point-of-sale devices, barcode readers
3. Healthcare - pharmaceutical records • Communications - telephone-call detail records
4. Location data - GPS, cell phones • Internet and e-commerce - Web logs, click-streams

3. Why Data Mining?

Handle explosive growth of data (from terabyte to petabyte) with the data mining techniques

Data / pattern Analysis in science and societal applications
(Ex: Market data analysis, weather data analysis, e-commerce applications, social data analysis, healthcare etc.)

Automated analysis of massive datasets

Evolution of Database Technology

1. Data Collection and Database Creation
2. Database Management System
3. Advanced Database Systems
4. Advanced Data Analysis (with OLAP tools in Data Warehouse)
5. Future Generation of Information Systems

Data Collection and Database Creation
(1960s and earlier)
■ Primitive file processing

Database Management Systems
(1970s to early 1980s)
■ Hierarchical and network database systems
■ Relational database systems
■ Data modeling: entity-relationship models, etc.
■ Indexing and accessing methods
■ Query languages: SQL, etc.
■ User interfaces, forms, and reports
■ Query processing and optimization
■ Transactions, concurrency control, and recovery
■ Online transaction processing (OLTP)

Advanced Database Systems
(mid-1980s to present)
■ Advanced data models: extended-relational, object relational, deductive, etc.
■ Managing complex data: spatial, temporal, multimedia, sequence and structured, scientific, engineering, moving objects, etc.
■ Data streams and cyber-physical data systems
■ Web-based databases (XML, semantic web)
■ Managing uncertain data and data cleaning
■ Integration of heterogeneous sources
■ Text database systems and integration with information retrieval
■ Extremely large data management
■ Database system tuning and adaptive systems
■ Advanced queries: ranking, skyline, etc.
■ Cloud computing and parallel data processing
■ Issues of data privacy and security

Advanced Data Analysis
(late-1980s to present)
■ Data warehouse and OLAP
■ Data mining and knowledge discovery: classification, clustering, outlier analysis, association and correlation, comparative summary, discrimination analysis, pattern discovery, trend and deviation analysis, etc.
■ Mining complex types of data: streams, sequence, text, spatial, temporal, multimedia, Web, networks, etc.
■ Data mining applications: business, society, retail, banking, telecommunications, science and engineering, blogs, daily life, etc.
■ Data mining and society: invisible data mining, privacy-preserving data mining, mining social and information networks, recommender systems, etc.

Future Generation of Information Systems
(Present to future)

Possible Questions



1. Define Data mining
2. What are the motivations of data mining?
3. What is the importance of data mining?
4. Describe the evolution of database technology
5. What is the need for data mining

Thank You