IT 308: Data Mining and Data Warehousing

(Elective)

Credits: 3
Lecture Hours: 48

Course Objective

The objective of the course is to make learner understand foundation principles and techniques of data mining and data warehousing. Students will be able to select and use various data mining language and tools very useful for adding business value of an organization.

Course Description

Introduction, Data Preprocessing- Data Integration and Transformation, Classification, Association Analysis, Cluster Analysis, Information Privacy and Data Mining, Advanced Applications, Search engines, Data Warehouses, Capacity Planning.

Course Details

Unit 1: Introduction **LH 2**

1.1. Data Mining Origin
1.2. Data Mining & Data Warehousing basics

Unit 2: Data Preprocessing **LH 6**

2.1. Data Types and Attributes
2.2. Data Pre-processing
2.3. OLAP
2.4 Characteristics of OLAP Systems
2.5 Multidimensional View and Data cube
2.6 Data Cube Implementation
2.7 Data Cube Operations
2.8 Guidelines for OLAP Implementation

Unit 3: Classification **LH 7**

3.1. Basics and Algorithms
3.2. Decision Tree Classifier
3.3. Rule Based Classifier
3.4. Nearest Neighbor Classifier
3.5. Bayesian Classifier
3.6. Artificial Neural Network Classifier
3.7. Issues : Overfitting, Validation, Model Comparison

Unit 4: Association Analysis **LH 7**

4.1. Basics and Algorithms
4.2. Frequent Itemset Pattern & *Apriori* Principle
4.3. FP-Growth, FP-Tree
4.4. Handling Categorical Attributes

Unit 5: Cluster Analysis **LH 7**

5.1. Basics and Algorithms
5.2. K-means Clustering
5.3. Hierarchical Clustering
5.4. DBSCAN Clustering

Unit 6: Information Privacy and Data Mining **LH 3**

6.1 Basic principles to Protect Information Privacy
6.2 Uses and Misuses of Data Mining
6.3 Primary Aims of data Mining
6.4 Pitfalls of Data Mining

Unit 7: Advanced Applications **LH 3**

7.1. Web-mining: Web content mining, web usage mining
7.2. Time-series data mining

Unit 8: Search Engines **LH 3**

8.1 Characteristics of search engine
8.2 Search Engine functionality
8.3 Ranking of Web pages

Unit 9: Data Warehousing **LH 7**

9.1 Operational Data sources
9.2 ETL (Extract, Transform, Load)
9.3 Data Warehouse Processes, Managers and their functions
9.4 Data Warehouses and Data Warehouses Design
9.5 Guidelines for Data Warehouse Implementation

Unit 10 Capacity Planning **LH 3**

10.1 Calculating storage requirement, CPU requirements

Practical:

Students should practice enough on real-world data intensive problems

References:

· Pang-NingTan, Michael Steinbach and Vipin Kumar, Introductionto Data Mining, 2005, Addison-Wesley.
· Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, 2nd Edition, 2006, Morgan Kaufmann.
- G.K. Gupta, Introduction to Data Mining with Case Studies, Prentice Hall of India
- IBM, An Introduction to Building the Data Warehouse, Prentice Hall of India
- IBM, Introduction to Business Intelligence and Data Warehousing, Prentice Hall of India
- Adriaans Pieter, D. Zantige, "Data Mining", Pearson Education Asia Pub. Ltd, 2002