

MCA 403: Data Warehousing and Data Mining

Teaching Scheme Lectures: 3 hrs/Week Tutorials: 1 hr/Week Credits: 4	Examination Scheme Class Test -12Marks Teachers Assessment - 6Marks Attendance – 12 Marks End Semester Exam – 70 marks
--	---

Prerequisite: -

1. Familiarity with the data base management system
2. Knowledge of repository system.

Course Objectives:

1. Be familiar with mathematical foundations of data mining tools.
2. Understand and implement classical models and algorithms in data warehouses and data mining
3. Characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering.
4. To explain the stages and process different data mining techniques. E. To learn mining and warehouse techniques through the use of different tools (e.g. ORACLE)

Detailed Syllabus

UNIT I (10 Hours)

Introduction: Fundamentals of data mining, Data Mining Functionalities, Classification of Data Mining systems, Major issues in Data Mining. **Data Pre-processing:** Needs, Pre-processing the Data, Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.

UNIT II (10 Hours)

Introduction: Data Warehouse and OLAP Technology for Data Mining, Data Warehouse Multidimensional Data Model, Data Warehouse Architecture, Data Warehouse Implementation, From Data Warehousing to Data Mining. Data Mining Primitives, Data Mining Query Languages.

UNIT III (10 Hours)

Concepts Description: Characterization and Comparison, Data Generalization and Summarization-Based Characterization. Analytical Characterization, Analysis of Attribute Relevance, Mining Class Comparisons: Discriminating between Different Classes, Mining Descriptive Statistical Measures in Large Databases.

UNIT IV (10 Hours)

Mining Association Rules in Databases: Association Rule Mining, Mining Single-Dimensional Boolean Association Rules from Transactional Databases, Mining Multilevel Association Rules from Transaction Databases, Mining Multidimensional Association Rules from Relational Databases and Data Warehouses, From Association Mining to Correlation Analysis.

UNIT V (6 Hours)

Classification and Prediction: Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Classification by Back-propagation, Classification Based on Association Rule Mining, Other Classification Methods, Prediction, and Classifier Accuracy.

UNIT VI (10 Hours)

Cluster Analysis Introduction: Types of Data in Cluster Analysis, a Categorization of Major Clustering Methods, Partitioning Methods, Density-Based Methods, Outlier Analysis. **Mining Complex Types of Data:** Multidimensional Analysis and Descriptive Mining of Complex, Data Objects, Mining-Spatial Databases, Multimedia Databases, Time-Series and Sequence Data, Text Databases, World Wide Web.

Text and Reference Books

1. Data Mining -Concepts and Techniques, Han, Kamber, Harcourt India, 2006.
2. Data Mining Introductory and advanced topics, Margaret H Dunham, Pearson, 2002.
3. Data Mining Techniques, Arjun K. Pujari, University Press, 2001.

Course Outcomes:

After completing the course, students will be able to:

1. The candidate will get knowledge of - Data preprocessing and data quality..
2. Modeling and design of data warehouses
3. Algorithms for data mining.
4. Be able to design data warehouses.
5. Ability to apply acquired knowledge for understanding data and select suitable methods for data analysis