

L	T	P	C
3	-	3	5

Course Description and Objectives:

This course is aimed at offering data and information management, information retrieval, and knowledge discovery in modern organizations. Case studies of those organizations using technologies to support business intelligence gathering and decision making are explored. This course is designed to understand the issues relating to the feasibility, usefulness, effectiveness, and scalability of techniques used for the discovery of patterns hidden in large data sets and also characterizes the kind of patterns that can be discovered by association rule mining, classification and clustering

Course Outcomes:

Students are able to

- ✓ Learn the basic concepts of Database Technology Evaluation steps and also understand the need of data mining and its functionalities
- ✓ Explore the efficient and effective maintenance of Data Warehouses.
- ✓ Apply the data mining functionalities like Clustering, Classification, Association Analysis to real world data.
- ✓ Discover interesting patterns and association rules from huge volume of data used to do classifications and predictions.
- ✓ Gain knowledge on developing areas like Web Mining, Text Mining, and Spatial Mining.

Skills:

- ✓ Design and development of schema models for a data warehouse
- ✓ Extraction of hidden interesting association rules
- ✓ Implementation of various classification and clustering algorithms
- ✓ Extraction of knowledge from text databases

UNIT- I

Introduction: Why Data Mining, What is Data Mining, Kinds of Data, Kinds of Patterns, and Technologies used, Kinds of applications adopted, Major issues in Data Mining.

Data Warehousing and Online Analytical Processing: Basic Concepts, Data Warehouse Modeling, Data Warehouse Design and Usage, Data Warehouse Implementation, Data Generalization by Attribute-Oriented Induction

UNIT- II

About Data: Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Measuring Data Similarity and Dissimilarity

Data Preprocessing: An Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization

UNIT- III

Data Cube Technology: Preliminary Concepts, Data Cube Computation Methods, Processing Advanced Kinds of Queries by Exploring Cube Technology, Multidimensional Data Analysis in Cube Space

Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods: Basic Concepts, Frequent Itemset Mining Methods, Which Patterns Are Interesting?—Pattern Evaluation Methods

Advanced Pattern Mining: Pattern Mining in Multilevel, Multidimensional Space, Constraint-Based Frequent Pattern Mining.

UNIT- IV

Classification: Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Model Evaluation and Selection, Techniques to Improve Classification Accuracy

Advanced Classification: Bayesian Belief Networks, Classification by Back propagation, Support Vector Machines, Classification Using Frequent Patterns, Lazy Learners, Other Classification Methods

UNIT- V

Cluster Analysis: Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Evaluation of Clustering

Advanced Cluster Analysis: Probabilistic Model-Based Clustering, Clustering High-Dimensional Data

LABORATORY EXPERIMENTS

Course Outcomes:

Students can able

- ✓ To evaluate the different models of OLAP and data preprocessing.
- ✓ To enlist various algorithms used in information analysis of Data Mining Techniques.
- ✓ To demonstrate the knowledge retrieved through solving problems

.List of Experiments

1. Explore various commands given in PL/SQL in Oracle 8.0
2. Execute multi-dimensional data model using SQL queries.
3. Implement various OLAP operations such as slice, dice, roll up, drill up, pivot etc.
4. Implementation of Text Mining on the data warehouse
5. Explore the correlation-ship analysis between the data set
6. Evaluate attribute relevance analysis on a weather data warehouse
7. Evaluate Information Gain of an attribute in the student database
8. Experiment to predict the class using the Bayesian classification
9. Find out a weight & bias updating using the Back Propagation Neural Network
10. To perform various data mining algorithms on the give data base using WEKA

TEXT BOOKS:

1. Jiawei Han, Micheline Kamber “ Data Mining: Concepts and Techniques” 3rd edition ,Morgan Kaufmann, 2012

REFERENCE BOOKS :

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, “Introduction to Data Mining”, First Edition,2012.
2. Ralph Kimball, Margy Ross, “The Data Warehouse Toolkit”, first edition John Wiley and Sons Inc., 2002.
3. Alex Berson, Stephen Smith, Kurt Thearling, “Building Data Mining Applications for CRM”, first edition, Tata McGraw Hill, 2000.
4. Margaret Dunham, “Data Mining: Introductory and Advanced Topics”, first edition, Prentice Hall, 2002.
5. Paulraj Ponnaiah, “Data Warehousing Fundamentals”, first edition, Wiley Publishers, 2001.