
Objectives:

By this course the students are able to:

1. Learn how data mining is a part of the natural evolution of database technology, characterize data mining as a process of knowledge discovery and know about the kinds of data objects and assess the similarities and dissimilarities between objects represented by various attribute types.
2. Know about the preprocessing techniques to improve the quality of mining.
3. Learn about Design and implementation of Data warehousing and On Line Analytical Processing (OLAP) which is an essential step in knowledge discovery process.
4. Learn about the computation, materialization and processing of queries by Cube technology.
5. Characterize the kinds of patterns that can be discovered by association rule mining.
6. Learn data classification techniques such as decision tree induction, Bayesian classification and Bayesian belief networks and techniques that builds classifier model for data analysis on large databases.
7. Know about the concepts of clustering and its techniques.
8. Identify the hidden outliers and know about detection methods.

Learning Outcomes:

At the end of the course the student will be able to:

1. Describe the evolution of database technology and characterize data mining as a process of knowledge discovery and assess the similarity and dissimilarity between objects.
2. Apply data preprocessing techniques such as data cleaning, data integration, data Transformation and data reduction strategies in order to improve the quality of the data, and mining results.
3. Design the warehouse schema to model a data warehouse that facilitates online analytical processing and the relationship between data warehouse and data mining.
4. Apply various cube operations and analyze queries on multidimensional data.
5. Determine interesting association or correlation relationships among a large set of data items using various association algorithms.
6. Apply data classification techniques such as decision tree induction, Bayesian classification and techniques that builds classifier model for data analysis on large databases.
7. Apply unsupervised learning techniques (Clustering) such as partitioning method, hierarchical method, density based method, grid based methods.
8. List various outliers by using outlier detection methods.

UNIT - I

Introduction: Data Mining, Importance of Data Mining, Kinds of Data, Kinds of Patterns, Technologies Used, Kinds of Applications, Major Issues in Data Mining, Data Objects and Attribute Types, Statistical Description of Data, Data Visualization, Measuring Data Similarity and Dissimilarity.

UNIT - II

Data Pre-processing: Major Tasks in Data Pre-processing, Data Cleaning and Data Integration, Data Reduction, Data Transformation and Data Discretization.

UNIT - III

Data Warehousing and On-Line Analytical Processing: Data Warehouse Modelling: Data Cube and OLAP, Data Warehouse Design and Usage, Data Warehouse Implementation, Data Generalization by Attribute-Oriented Induction.

UNIT - IV

Data Cube Technology: Cube Materialization, Cube Computation, Processing Advanced kinds of Queries by Exploring Cube Technology, Multidimensional Data Analysis in Cube Space.

UNIT - V

Mining Frequent Patterns, Associations and Correlations: Basic Concepts, Frequent Item set Mining Methods, Pattern Interestingness and Evaluation Methods.
Advanced Pattern Mining: Pattern Mining in Multilevel, Multi Dimensional Space, Constraint-Based Frequent Pattern Mining, Mining High Dimensional Data and Colossal Patterns, Mining Compressed or Approximate Patterns, Pattern Exploration and Application.

UNIT - VI

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

Classification Advanced Methods: Bayesian Belief Networks, Classification by Back Propagation, Support Vector Machines, Classification Using Frequent Patterns, Lazy Learners, Other Classification Methods, Additional Topics Regarding Classification.

UNIT - VII

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

Advanced Cluster Analysis: Probabilistic Model-Based Clustering, Clustering High-Dimensional Data, Clustering Graph and Network Data, Clustering with Constraints.

UNIT - VIII

Outlier Analysis: Outliers, Outlier Detection Methods, Statistical Approaches, Proximity-Based Approaches, Clustering Based Approaches, Classification Based Approaches, Mining Contextual and Collective Outliers, Outlier Detection in High-Dimensional Data, Mining Complex Data Types, Other Methodologies of Data Mining.

Learning Resources

Text Book:

1. Data Mining, Concepts and Techniques, JIAWEI HAN, MICHELINE KAMBER & JIAN PEI, Third Edition, Morgan Kaufmann Publications 2012.

Reference Books:

1. Data Mining Techniques, ARUN K PUJARI, University Press.
2. Data Warehousing in the Real World, SAM ANAHORY & DENNIS MURRAY, Pearson Education.
3. Data Warehouse Design, MATTEO GOLFARELLI & STEFANO RIZZI, Tata McGraw Hill Edition.
4. Data Mining Introductory and Advanced topics, MARGARET H DUNHAM, Pearson Education.