

DATA MINING

OBJECTIVES:

- Students will be enabled to understand and implement classical models and algorithms in data warehousing and data mining.
- They will learn how to analyze the data, identify the problems, and choose the relevant models and algorithms to apply.
- They will further be able to assess the strengths and weaknesses of various methods and algorithms and to analyze their behavior.

UNIT –I:

Introduction: Why Data Mining? What Is Data Mining? 1.3 What Kinds of Data Can Be Mined? 1.4 What Kinds of Patterns Can Be Mined? Which Technologies Are Used? Which Kinds of Applications Are Targeted? Major Issues in Data Mining. Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Measuring Data Similarity and Dissimilarity

UNIT –II:

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

UNIT –III:

Classification: Basic Concepts, General Approach to solving a classification problem, Decision Tree Induction: Working of Decision Tree, building a decision tree, methods for expressing an attribute test conditions, measures for selecting the best split, Algorithm for decision tree induction.

UNIT –IV:

Classification: Alternative Techniques, Bayes' Theorem, Naïve Bayesian Classification, Bayesian Belief Networks

UNIT –V

Association Analysis: Basic Concepts and Algorithms: Problem Defecation, Frequent Item Set generation, Rule generation, compact representation of frequent item sets, FP-Growth Algorithm. (Tan & Vipin)

UNIT –VI

Cluster Analysis: Basic Concepts and Algorithms: Overview: What Is Cluster Analysis? Different Types of Clustering, Different Types of Clusters; K-means: The Basic K-means Algorithm, K-means Additional Issues, Bisecting K-means, Strengths and Weaknesses; Agglomerative Hierarchical Clustering: Basic Agglomerative Hierarchical Clustering Algorithm DBSCAN: Traditional Density Center-Based Approach, DBSCAN Algorithm, Strengths and Weaknesses. (**Tan & Vipin**)

OUTCOMES:

- Understand stages in building a Data Warehouse
- Understand the need and importance of preprocessing techniques
- Understand the need and importance of Similarity and dissimilarity techniques
- Analyze and evaluate performance of algorithms for Association Rules.
- Analyze Classification and Clustering algorithms

TEXT BOOKS:

1. Introduction to Data Mining: Pang-Ning Tan & Michael Steinbach, Vipin Kumar, Pearson.
2. Data Mining concepts and Techniques, 3/e, Jiawei Han, Michel Kamber, Elsevier.

REFERENCE BOOKS:

1. Data Mining Techniques and Applications: An Introduction, Hongbo Du, Cengage Learning.
2. Data Mining : Vikram Pudi and P. Radha Krishna, Oxford.
3. Data Mining and Analysis - Fundamental Concepts and Algorithms; Mohammed J. Zaki, Wagner Meira, Jr, Oxford
4. Data Warehousing Data Mining & OLAP, Alex Berson, Stephen Smith, TMH.