

DATA WAREHOUSING AND MINING**UNIT-1:**

Introduction to Data mining, types of Data, Data Quality, Data Processing, Measures of Similarity and Dissimilarity, Exploring Data: Data Set, Summary Statistics, Visualization, OLAP and multi dimensional data analysis.

UNIT-II:

Classification: Basic Concepts, Decision Trees and model evaluation: General approach for solving a classification problem, Decision Tree induction, Model over fitting: due to presence of noise, due to lack of representation samples, Evaluating the performance of classifier. Nearest Neighborhood classifier, Bayesian Classifier, Support vector Machines: Linear SVM, Separable and Non Separable case.

UNIT-III:

Association Analysis: Problem Definition, Frequent Item-set generation, rule generation, compact representation of frequent item sets, FP-Growth Algorithms. Handling Categorical, Continuous attributes, Concept hierarchy, Sequential, Sub graph patterns

UNIT-IV:

Clustering: Over view, K-means, Agglomerative Hierarchical clustering, DBSCAN, Cluster evaluation: overview, Unsupervised Cluster Evaluation using cohesion and separation, using proximity matrix, Scalable Clustering algorithm

UNIT-V:

Web data mining: Introduction, Web terminology and characteristics, Web content mining, Web usage mining, web structure mining,
Search Engines: Characteristics, Functionality, Architecture, Ranking of WebPages, Enterprise search

TEXTBOOKS:

1. Introduction to Data Mining: Pang-Ning tan, Michael Steinbach, Vipin kumar, Addison-Wesley.
2. Introduction to Data Mining with Case Studies: GK Gupta; Prentice Hall.

REFERENCEBOOKS:

1. Data Mining: Introductory and Advanced Topics, Margaret H Dunham, Pearson, 2008.
2. Fundamentals of data warehouses, 2/e, Jarke, Lenzerini, Vassiliou, Vassiliadis, Springer.
3. Data Mining Theory and Practice, Soman, Diwakar, Ajay, PHI, 2006.
4. Data Mining, Concepts and Techniques, 2/e, Jiawei Han, Micheline Kamber, Elsevier, 2006.