**Data Warehousing & Mining**

**UNIT- I**

**Data Warehouse and OLAP Technology:** An Overview: Data Warehouse, A Multidimensional Data Model, Data Warehouse Architecture, Data Warehouse Implementation, From Data Warehousing to Data Mining. (Han & Kamber)

**UNIT- II**

**Data Mining:** Introduction, Data Mining, Motivating challenges, The origins of Data Mining, Data Mining Tasks, Types of Data, Data Quality.

Data Preprocessing: Aggregation, Sampling, Dimensionality Reduction, Feature Subset Selection, Feature creation, Discretization and Binarization, Variable Transformation, Measures of Similarity and Dissimilarity. (Tan &Vipin)

**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.

**Model Overfitting:** Due to presence of noise, due to lack of representation samples, evaluating the performance of classifier: holdout method, random sub sampling, cross-validation, bootstrap. Bayes Theorem, Naïve Bayes Classifier (Tan &Vipin)

**UNIT -IV**

**Association Analysis:** Basic Concepts and Algorithms: Problem Definition, Frequent Item Set Generation, Apriori Principle, Apriori Algorithm, Rule Generation, Compact Representation of Frequent Itemsets, FPGrowth Algorithm. (Tan &Vipin)

**UNIT –V**

**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)