

**DS7210E: DATA MINING (Open Elective – II)****B.Tech. IV Year I Sem.**

L	T	P	C
3	0	0	3

**Pre-Requisites:**

1. Database Management System
2. Probability and Statistics

**Course Objectives:**

- Students will become acquainted with both the strengths and limitations of various data mining techniques like Association, Classification, Cluster and Outlier analysis.

**Course Outcomes:**

- Understand the need of data mining and pre-processing techniques.
- Perform market basket analysis using association rule mining.
- Utilize classification techniques for analysis and interpretation of data.
- Identify appropriate clustering and outlier detection techniques to handle complex data.
- Understand the mining of data from web, text and time series data.

**UNIT - I****Introduction to Data Mining:**

What Data mining? Kinds of Data, Knowledge Discovery process, Data Mining Functionalities, Kinds of Patterns, Major Issues in Data Mining. Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Measuring Data Similarity and Dissimilarity, Data Pre-processing: Major Tasks in Data Pre-processing, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization.

**UNIT - II**

**Association Analysis:** Basic Concepts, Market Basket Analysis, Apriori Algorithm, FP-growth, From Association Analysis to Correlation Analysis, Pattern Mining in Multilevel Associations and Multidimensional Associations.

**UNIT - III**

**Classification:** Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Metrics for Evaluating Classifier Performance, Ensemble Methods, Multilayer Feed-Forward Neural Network, Support Vector Machines, k-Nearest-Neighbor Classifiers.

**UNIT - IV**

**Cluster Analysis:** Requirements for Cluster Analysis, Overview of Basic Clustering Methods, Partitioning Methods-k-Means, k-Medoids, Hierarchical Methods-AGENES, DIANA, BIRCH, Density-Based Method-DBSCAN, Outlier Analysis: Types of Outliers, Challenges of Outlier Detection, and Overview of Outlier Detection Methods

**UNIT - V**

**Advanced Concepts:** Web Mining- Web Content Mining, Web Structure Mining, Web Usage Mining, Spatial Mining- Spatial Data Overview, Spatial Data Mining Primitives, Spatial Rules, Spatial Classification Algorithm, Spatial Clustering Algorithms, Temporal Mining- Modeling Temporal Events, Time Series, Pattern Detection, Sequences, Temporal Association Rules.

**TEXT BOOKS:**

1. Jiawei Han, Micheline Kamber, Jian Pei., Data Mining: Concepts and Techniques, 3<sup>rd</sup> Edition, Morgan Kaufmann/Elsevier, 2012.
2. Margaret H Dunham, Data Mining Introductory and Advanced Topics, 2<sup>nd</sup> Edition, Pearson Education, India, 2006.

**REFERENCE BOOKS:**

1. Data Mining Techniques, Arun K Pujari, 3<sup>rd</sup> Edition, Universities Press.
2. Pang-Ning Tan, Michael Steinbach, Anuj Karpatne and Vipin Kumar, Introduction to Data Mining, 2<sup>nd</sup> Edition, Pearson Education India, 2021.
3. Amitesh Sinha, Data Warehousing, Thomson Learning, India, 2007.