**4/4 B.Tech. SEVENTHSEMESTER**
**ELECTIVE – III**

**EM7T4B DATA WARE HOUSING AND DATA MINING Credits: 3**

**Lecture: 3 periods/week** **Internal assessment: 30 marks**
**Tutorial: 1 period /week** **Semester end examination: 70 marks**

---

**Objectives:**

Giving different techniques to extract knowledge and different visualization techniques

**Outcomes:**

After completion of the course Students will be able to

- Understand the preprocessing techniques and OLAP operations
- Able to write DMQL
- Depends on the problem able to apply appropriate data mining algorithm.

**UNIT I:**
**Data Mining**: Introduction, Data Mining, Kinds of Data, Data Mining Functionalities, Classification of Data Mining Systems, Major issues in Data Mining. A Multi-dimensional data model, Data Warehouse Architecture, Data Warehouse Implementation.

**UNIT II:**
**Data Preprocessing**: Data cleaning, Data Integration & Transformation, Data Reduction, Discretization & Concept Hierarchy Generation, Data Mining Primitives.

**Unit III:**
**Data Analysis Techniques Online Analytical Processing :** OLAP, differences between OLAP and OLTP systems, Multi Dimensional Data Model , OLAP operators, Relational DBMS support for OLAP, Data Cube Demonstration using SQL , Various Categories of OLAP Tools , Efficient processing of OLAP queries

**Unit IV:**
**Data Mining Primitives**, Languages and system Architectures, Data Mining Primitives: What defines a Data Mining Task?, A Data Mining query language(DMQL) , Designing Graphical User Interfaces Based on a Data Mining Query language, Architectures of Data Mining Systems

**Unit V:**
**Mining Association rules in Large Databases:** Association rule mining, mining single-dimensional, Boolean Association rules from Transactional Databases, Mining Multi-dimensional Association rules from relational databases & Data Warehouses.

**UNIT VI:**
**Classification and Prediction:** Introduction, Classification by Decision tree induction, Bayesian Classification, Classification by Back propagation, Other Classification Methods, Prediction, Classifier accuracy.

**UNIT VII:**

**Cluster Analysis**: Introduction, Types of data in Cluster analysis, A categorization of major clustering methods, partitioning methods, Hierarchical methods, Density-Based Methods: DBSCAN, Grid-based Method: STING; Model-based Clustering Method: Statistical approach, Outlier analysis.

**Unit VIII:**

**Mining Stream, Time-Series:** Sequence Data, Graph Mining, and Multi relational Data Mining, Spatial, Multimedia, Text, and Web Data Applications and Trends in Data Mining.

**TEXT BOOKS:**

1. Jiawei Han Micheline Kamber, Data Mining Concepts and Techniques. Morgan Kaufmann Publishers.
2. Introduction to Data Mining – Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Pearson Education.

**REFERENCE BOOKS**:

1. Ralph Kinball, Data Warehouse Toolkit. John Wiley Publishers.
2. Margaret H.Dunham, Data Mining (Introductory and Advanced Topics). Pearson Education.
3. Sam Anahory, Dennis Murray, Data Warehousing in the Real World – A Practical Guide for BuildingDecision Support Systems. Pearson Education.
4. G.K.Gupta, Introduction to Data Mining with Case Studies. PHI, 2006.