III/IV B. TECH. SECOND SEMESTER
DATA WAREHOUSING AND DATA MINING(Required)

Course Code:CS 6T4   **Credits: 3**
Lecture: 3 periods/ week   **Internal assessment: 30 Marks**
Tutorial: 1period/week   **Semester end examination: 70 Marks**

Prerequisite: Database Management Systems

Course Objectives:

By this course the student will be able to:

1. Understand data mining as a process of knowledge discovery and also about the pre-processing techniques to improve the quality of mining.
2. Learn about Data warehousing and On Line Analytical Processing (OLAP).
3. Understand the kinds of patterns that can be discovered by association rule mining and different classification techniques that builds classifier model for data analysis.
4. Know about the basic concepts of clustering.
5. Evaluate methodological issues underlying the effective application of data mining.

Course Outcomes:

At the end of this course student will:

CO1) Understand the fundamentals of data mining and data warehousing concepts

CO2) Explain the processing of raw input data for data mining applications

CO3) Discover the interesting patterns from different kinds of databases

CO4) Demonstrate supervised (classification) and unsupervised (clustering) learning techniques

CO5) Describe the outlier detection methods and various data mining methodologies for complex data types

Syllabus:

UNIT 1
Data Warehousing and Online Analytical Processing: Data Warehouse**:** Basic Concepts, Data Warehouse Modeling: Data Cube and OLAP, Data Warehouse Design and Usage, Data Warehouse Implementation.

UNIT 2

Introduction: Fundamentals of data mining: Kinds of data, Data Mining Functionalities, Classification of Data Mining systems, Major issues in Data Mining.

Data Preprocessing: Need for Preprocessing the Data, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Discretization.

UNIT 3

Mining Frequent Patterns, Associations, and Correlations: Basic Concepts, Frequent Item Set Mining Methods.

Classification: Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule Based Classification.

UNIT 4

Cluster Analysis: Basic Concepts and Methods, Cluster Analysis, **Partitioning Methods:** k-means and k-mediods**, Hierarchical Method:** Agglomerative Hierarchical clustering (BIRCH**), Density-Based Methods**: DBSCAN, **Grid-based Methods**.

UNIT 5

Outlier Detection: Outliers and Outlier Analysis, Outlier Detection Methods. **Data Mining Trends:** Mining Complex Data Types, Other Methodologies of Data Mining.

Learning Resource

Text Books

1. Data Mining – Concepts and Techniques – 3/e, Jiawei Han , Micheline Kamber & Jian Pei- Elsevier.

References

1. Introduction to Data Mining with Case Studies – 2nd Edition, G.K.Gupta, PHI
2. Introduction to Data Mining: Pang-Ning Tan, Michael Steinbach, Vipin Kumar,Pearson.
3. Data Mining Techniques – ARUN K PUJARI, University Press.
4. Data Warehousing in the Real World, SAM ANAHORY & DENNIS MURRAY, Pearson Edn. Asia.
5. Data Warehousing Fundamentals, PAULRAJ PONNAIAH WILEY STUDENT EDITION.