3/4 B.Tech. FIRST SEMESTER

IT5T1 DATA WAREHOUSING AND DATA MINING Credits: 4

Lecture: 4 periods/week **Internal assessment: 30 marks**
Tutorial: 1 period /week **Semester end examination: 70 marks**

Objectives:

- To give students a good overview of the ideas and the techniques behind recent developments in the data warehousing.
- To compare and contrast different conceptions of data mining as evidenced in real world applications.
- To explain the role of finding associations in commercial market basket data.
- To characterize the kinds of patterns that can be discovered by association rule mining.
- To describe how to extend a relational system to find patterns using association rules.
- To evaluate methodological issues underlying the effective application of data mining.
- To identify and characterize sources of noise, redundancy, and outliers in presented data.

Outcomes:

Students will be able to

- Understand the basic principles of Data and Knowledge Mining.
- Understand the concepts of data warehousing and OLAP.
- Apply the Data and Knowledge Mining tools to various business problems.
- Select and use appropriate software for business applications.
- Demonstrate the knowledge gained through solving problems particularly in the concerned laboratory course.

Syllabus:

UNIT – I
INTRODUCTION:
Fundamentals of data mining, Data Mining Functionalities, Classification of Data Mining systems, Major issues in Data Mining.
DATA PREPROCESSING:
Needs Preprocessing the Data, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Discretization**. (Chapters-1, 3)**

UNIT – II
DATA WAREHOUSING AND ONLINE ANALYTICAL PROCESSING:
Basic Concepts, Data Warehouse Modeling, Data Cube and OLAP, Data Warehouse Implementation, Data Generalization by Attribute-Oriented Induction (AOI). **(Chapter-4)**

UNIT – III
Data Objects and Attribute Types, Basic Statistical Description of Data, Measuring Data Similarity and Dissimilarity**. (Chapter 2)**

UNIT – IV
MINING FREQUENT PATTERNS, ASSOCIATIONS AND CORRELATIONS:
Basic Concepts, Frequent Item set Mining Methods, Pattern Evaluation Methods and Pattern Mining in Multilevel, Multidimensional Space**. (Chapters-6, 7)**

UNIT – V
CLASSIFICATION:
Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Model Evaluation and Selection, Techniques to Improve Classification Accuracy**. (Chapter-8)**

UNIT – VI
CLUSTER ANALYSIS:
Basic Concepts and Methods, Cluster Analysis, Partitioning Methods, Hierarchical Methods**. Chapter -10)**

UNIT – VII
CLUSTER ANALYSIS:
Density-Based Methods, Grid-Based Methods, Evaluation of Clustering.
Outlier Detection:
Outliers and Outlier Analysis, Outlier Detection Methods**. (Chapters-10, 12)**

UNIT – VIII
DATA MINING TRENDS:
Mining Complex Data Types, Other Methodologies of Data Mining, Data Mining Applications, Data Mining and Society.
(Chapter-13)

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

Reference Books:
1. Introduction to Data Mining Pang-Ning Tan, Michael Steinbach, Vipin Kumar,Pearson
2. Data Mining Techniques – Arun K Pujari, University Press.
3. Data Warehousing in the Real World – Sam Anahory & Dennis Murray. Pearson Edn Asia.
4 Data Warehousing Fundamentals – Paulraj Ponnaiah Wiley Student Edition.
5. The Data Warehouse Life cycle Tool kit – Ralph Kimball Wiley Student Edition.