

**3/4 B.Tech. SECOND SEMESTER
DATA WAREHOUSING AND DATA MINING**

CS6T3

Required

Credits: 4

Lecture: 4 periods/week

Internal assessment: 30 marks

Tutorial: 1 period /week

Semester end examination: 70 marks

Course context and Overview: This course gives an introduction to methods and theory for development of data warehouses and data analysis using data mining. Data quality and methods and techniques for preprocessing of data. Modeling and design of data warehouses. Algorithms for classification, clustering and association rule analysis. Practical use of software for data analysis.

Prerequisites: Algorithms, Data structures, Probability, Statistics, DBMS.

Objectives:

By this course the students will be able to:

1. Understand data mining as a process of knowledge discovery and also about the preprocessing techniques to improve the quality of mining.
2. Learn about Data warehousing and On Line Analytical Processing (OLAP).
3. Know about the kinds of data objects, attribute types and can assess the similarities and dissimilarities between objects.
4. Understand the kinds of patterns that can be discovered by association rule mining.
5. Learn about different classification techniques that builds classifier model for data analysis.
6. Know about the basic concepts of clustering and its techniques.
7. Identify the hidden outliers and know about detection methods.
8. Evaluate methodological issues underlying the effective application of data mining.

Learning Outcomes:

At the end of the course the student should be able to:

Ability to:

1. Understand the fundamentals of data mining and data warehousing concepts.
2. Evaluate raw input data and process it for data mining applications.
3. Discover the interesting patterns from different kinds of databases.
4. Demonstrate supervised (classification) and unsupervised (clustering) learning techniques.
5. Describe the outlier detection methods.
6. Discuss various data mining methodologies and applications including societal relevance.

UNIT – I

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. (**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 Itemset 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**)

Learning Resources

TEXT BOOK:

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.