3/4 B.Tech - SECOND SEMESTER

IT6T4   **DATA MINING AND DATA WAREHOUSING**   **Credits: 3**

Lecture: 3 Periods/week   **Internal assessment: 30 marks**

Practice/Interaction: 1Period/week   **Semester end examination: 70 marks**

Objectives:
- To provide an overview of the techniques and developments in the data warehousing and mining.
- To explain the role of data warehousing techniques and applicability in commercial data.
- To characterize the kinds of patterns using association rule mining and classification.
- To introduce basic concepts of clustering and outliers present in data.

Outcomes:

Students will be able to
- Understand the basic principles of Data Mining and data preprocessing.
- Differentiate the concepts of data warehousing and OLTP.
- Relate the learned algorithms in association and pattern mining to the practical issues.
- Describe and utilize a range of techniques for classifying the data and accuracy improvements.
- Analyze the data and develop some clustering and outlier methods.

Prerequisite:

Database Systems

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.

UNIT – II

Data Warehousing and Online Analytical Processing: Basic Concepts, Data Warehouse Modeling: Data Cube and OLAP. Data Objects and Attribute Types, Basic Statistical Description of Data, Measuring Data Similarity and Dissimilarity.

UNIT – III

Mining Frequent Patterns, Associations, and Correlations: Basic Concepts, Frequent Item set Mining Methods, Pattern Evaluation Methods, and Pattern Mining in Multilevel, Multidimensional Space.

UNIT-IV

Classification: Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Model Evaluation and Selection, Techniques to Improve Classification Accuracy.

UNIT – V

Cluster Analysis: Basic Concepts and Methods, Cluster Analysis, Partitioning Methods, Hierarchical Methods. Cluster Analysis: Density-Based Methods, Grid-Based Methods, Evaluation of Clustering. Outlier Detection: Outliers and Outlier Analysis, outlier Detection Methods. Introduction to text mining.

Text Book:

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

Reference Books:

1. Introduction to Data Mining: Pang-Ning Tan, Michael Steinbach, VipinKumar, 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 – PaulrajPonnaiah Wiley Student Edition.
5. The Data Warehouse Life cycle Tool kit – Ralph Kimball Wiley Student Edition.

e-Learning Resources:

1. https://weka.waikato.ac.nz/explorer
2. http://rapidminerresources.com
3. https://www.coursera.org