

## **CS8075 DATA WAREHOUSING AND DATA MINING**

### DETAILED SYLLABUS

#### **OBJECTIVES:**

- To understand data warehouse concepts, architecture, business analysis and tools
- To understand data pre-processing and data visualization techniques
- To study algorithms for finding hidden and interesting patterns in data
- To understand and apply various classification and clustering techniques using tools.

#### **UNIT I DATA WAREHOUSING, BUSINESS ANALYSIS AND ON-LINE ANALYTICAL PROCESSING (OLAP)**

Basic Concepts - Data Warehousing Components – Building a Data Warehouse – Database Architectures for Parallel Processing – Parallel DBMS Vendors - Multidimensional Data Model – Data Warehouse Schemas for Decision Support, Concept Hierarchies -Characteristics of OLAP Systems – Typical OLAP Operations, OLAP and OLTP.

#### **UNIT II DATA MINING – INTRODUCTION**

Introduction to Data Mining Systems – Knowledge Discovery Process – Data Mining Techniques – Issues – applications- Data Objects and attribute types, Statistical description of data, Data Preprocessing – Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures.

#### **UNIT III DATA MINING - FREQUENT PATTERN ANALYSIS**

Mining Frequent Patterns, Associations and Correlations – Mining Methods- Pattern Evaluation Method – Pattern Mining in Multilevel, Multi-Dimensional Space – Constraint Based Frequent Pattern Mining, Classification using Frequent Patterns

#### **UNIT IV CLASSIFICATION AND CLUSTERING**

Decision Tree Induction - Bayesian Classification – Rule Based Classification – Classification by Back Propagation – Support Vector Machines — Lazy Learners – Model Evaluation and Selection-Techniques to improve Classification Accuracy. Clustering Techniques – Cluster analysis-Partitioning Methods - Hierarchical Methods – Density Based Methods - Grid Based Methods – Evaluation of clustering – Clustering high dimensional data- Clustering with constraints, Outlier analysis-outlier detection methods.

#### **UNIT V WEKA TOOL**

Datasets – Introduction, Iris plants database, Breast cancer database, Auto imports database - Introduction to WEKA, The Explorer – Getting started, Exploring the explorer, Learning algorithms, Clustering algorithms, Association–rule learners.

#### **TEXT BOOK:**

1. Jiawei Han and Micheline Kamber, —Data Mining Concepts and TechniquesII, Third Edition, Elsevier, 2012.

**REFERENCES:**

1. Alex Berson and Stephen J. Smith, —Data Warehousing, Data Mining & OLAPII, Tata McGraw – Hill Edition, 35th Reprint 2016.
2. K.P. Soman, Shyam Diwakar and V. Ajay, —Insight into Data Mining Theory and Practicell, Eastern Economy Edition, Prentice Hall of India, 2006.
3. Ian H. Witten and Eibe Frank, —Data Mining: Practical Machine Learning Tools and TechniquesII, Elsevier, Second Edition.