

# IT 274: Data Warehousing and Data Mining

*Credits: 3*

*Lecture Hours: 48*

## **Course Objectives:**

The main objective of this course is to provide knowledge of different concepts of data warehousing and data mining to students. After completing this course, students will be able to

- Understand different concepts of data warehousing and data mining.
- Learn to design and use data warehouses and OLAP.
- Know and apply pattern mining algorithms.
- Understand and use classification and clustering algorithms.
- Learn outlier detection and related algorithms.

## **Course Description:**

This course covers different concepts of data warehousing and data mining including data mining, data warehousing and OLAP, and different concepts and algorithms of pattern mining, classification, clustering and outlier detection.

## **Course Details**

### **Unit 1: Introduction**

**6 LHs**

Data Mining and Knowledge Discovery; Mining Various Kinds of Knowledge; Data Mining Applications; Data Types; Statistics of Data; Similarity and Distance Measures; Data Quality, Data Cleaning, and Data Integration; Data Transformation; Dimensionality Reduction

### **Unit 2: Data Warehousing and Online Analytical Processing**

**8 LHs**

Data Warehouse; Data Warehouse Modeling: Schema and Measures; OLAP Operations; Data Cube Computation; Data Cube Computation Methods

### **Unit 3: Pattern Mining**

**5 LHs**

Basic Concepts; Frequent Itemsets, Closed Itemsets, and Association Rules; Frequent Itemset Mining Methods: Apriori and Pattern-Growth Approach; Generating Association Rules from Frequent Itemsets; Pattern Evaluation Methods

### **Unit 4: Classification**

**12 LHs**

Basic Concepts; Decision Tree Induction; Bayes Classification Methods; Lazy Learners; Linear Classifiers; Model Evaluation and Selection; Techniques to Improve Classification Accuracy

### **Unit 5: Cluster Analysis**

**10 LHs**

Cluster Analysis; Partitioning Methods (k-Means, k-Medoids and k-Modes); Hierarchical Methods (Agglomerative and Divisive); Density-Based (DBSCAN); Evaluation of Clustering

**Unit 6: Outlier Detection****7 LHs**

Basic Concepts; Statistical Approaches; Proximity-Based Approaches; Reconstruction-Based Approaches; Clustering-Based Approaches; Classification-Based Approaches

**Laboratory Works:**

The laboratory work includes implementing data warehouse and algorithms for pattern-mining, classification, clustering and outlier detection using any appropriate programming language. Students should also learn to use visualization tools during implementation of these algorithms.

**Suggested Reading**

1. Jiawei Han, Jian Pei, and Hanghang Tong, Data Mining Concepts and Techniques, 4<sup>th</sup> Edition, Morgan Kaufmann, 2023.