

IV Year - I Semester

<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
<b>4</b>	<b>0</b>	<b>0</b>	<b>3</b>

## DATA WAREHOUSING AND BUSINESS INTELLIGENCE

### OBJECTIVES:

- Approach business problems data-analytically by identifying opportunities to derive business value from data.
- Know the basics of data mining techniques and how they can be applied to extract relevant Business in

### UNIT- I:

Introduction to Data Mining: Motivation for Data Mining, Data Mining-Definition & Functionalities, Classification of DM systems, DM task primitives, Integration of a Data Mining system with a Database or a Data Warehouse, Major issues in Data Mining. **Data Warehousing (Overview Only):** Overview of concepts like star schema, fact and dimension tables, OLAP operations, From OLAP to Data Mining.

### UNIT -II:

Data Preprocessing: Why? Descriptive Data Summarization, Data Cleaning: Missing Values, Noisy Data, Data Integration and Transformation. Data Reduction:-Data Cube Aggregation, Dimensionality reduction, Data Compression, Numerosity Reduction, Data Discretization and Concept hierarchy generation for numerical and categorical data.

### UNIT- III:

Mining Frequent Patterns, Associations, and Correlations: Market Basket Analysis, Frequent Itemsets, Closed Itemsets, and Association Rules, Frequent Pattern Mining, Efficient and Scalable Frequent Itemset Mining Methods, The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation, Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, Frequent Itemsets without Candidate Generation using FP Tree, Mining Multilevel Association Rules, Mining Multidimensional Association Rules, From Association Mining to Correlation Analysis, Constraint-Based Association Mining.

### UNIT- IV:

Classification & Prediction: What is it? Issues regarding Classification and prediction  
**Classification methods:** Decision tree, Bayesian Classification, Rule based Prediction: Linear and non linear regression, Accuracy and Error measures, Evaluating the accuracy of a Classifier or Predictor.

**Cluster Analysis:** What is it? Types of Data in cluster analysis, Categories of clustering methods, Partitioning methods ñ K-Means, K-Medoids. Hierarchical Clustering- Agglomerative and Divisive Clustering, BIRCH and ROCK methods, DBSCAN, Outlier Analysis

#### **UNIT- V:**

**Mining Stream and Sequence Data:** What is stream data? Classification, Clustering Association Mining in stream data. Mining Sequence Patterns in Transactional Databases.  
**Spatial Data and Text Mining:** Spatial Data Cube Construction and Spatial OLAP, Mining Spatial Association and Co-location Patterns, Spatial Clustering Methods, Spatial Classification and Spatial Trend Analysis. Text Mining Text Data Analysis and Information Retrieval, Dimensionality Reduction for Text, Text Mining Approaches.

#### **UNIT- VI:**

**Web Mining:** Web mining introduction, Web Content Mining, Web Structure Mining, Web Usage mining, Automatic Classification of web Documents.

**Data Mining for Business Intelligence Applications:** Data mining for business Applications like Balanced Scorecard, Fraud Detection, Click stream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc

#### **OUTCOMES**

- Describe the scope and application of business intelligence and decision support;
- Design systems for sourcing and structuring data to provide an integrated, non-volatile collection of data for decision support using data warehouses;
- Design multidimensional data models and implement them using star schemas and relational databases;
- Communicate and foster realistic expectations of the role of OLAP technology and business intelligence systems in management and decision support;
- Explain the need for evolutionary development approaches to developing business intelligence and data warehouse systems;
- Develop a simple business intelligence system using an OLAP tool;
- Apply theories and principles of data visualization to encourage high quality analysis of business information to inform decision making;
- Design governance mechanisms for the development and management of business intelligence and data warehouse systems in an organization.

#### **TEXT BOOKS:**

1. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 2 nd Edition
2. P. N. Tan, M. Steinbach, Vipin Kumar, introduction to Data Mining, Pearson Education

## REFERENCE BOOKS:

1. MacLennan Jamie, Tang ZhaoHui and Crivat Bogdan, Data Mining with Microsoft SQL Server 2008, Wiley India Edition.
2. G. Shmueli, N.R. Patel, P.C. Bruce, Data Mining for Business Intelligence: Concepts, Techniques and Applications in Microsoft Office Excel with XLMiner, Wiley India.
3. Michael Berry and Gordon Linoff Data Mining Techniques, 2nd Edition Wiley Publications.
4. Alex Berson and Smith, Data Mining and Data Warehousing and OLAP, McGraw Hill Publication.
5. E. G. Mallach, Decision Support and Data Warehouse Systems", Tata McGraw Hill.
6. Michael Berry and Gordon Linoff Mastering Data Mining- Art & science of CRM, Wiley Student Edition
7. Arijay Chaudhry & P. S. Deshpande, Multidimensional Data Analysis and Data Mining Dreamtech Press
8. Vikram Pudi & Radha Krishna, Data Mining, Oxford Higher Education.
9. Chakrabarti, S., Mining the Web: Discovering knowledge from hypertext data,
10. M. Jarke, M. Lenzerini, Y. Vassiliou, P. Vassiliadis (ed.), Fundamentals of Data Warehouses, Springer-Verlag, 1999.