**Bachelor of Technology (Computer Science and Engineering)**

**Semester-V**

L-2 T-1 P-0 C-3

**CSE776 TR1: Data Warehousing & Mining**

**Course Objectives**

* To learn the principles of Data warehouses and Data Mining
* To be familiar with the Data warehouse architecture and its Implementation
* To know the Architecture of a Data Mining system
* To learn the various Data preprocessing Methods
* To perform classification and prediction of data

**Course Outcomes (COs):** Upon completion of this unit students will be able to:

1. Understand the functionality of the various data mining and data warehousing component
2. Analyze the strengths and limitations of various data mining and data warehousing models
3. analyzing techniques of various data mining
4. Understand different methodologies used in data mining and data ware housing.
5. Analyze the different approaches of data ware housing and data mining with various technologies

**Articulation Matrix**

*(Program Articulation Matrix is formed by the strength of correlation of COs with POs and PSOs. The strength of correlation is indicated as 3 for substantial (high), 2 for moderate (medium) correlation, and 1 for slight (low) correlation)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO/PO/PSO** | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** | **PSO1** | **PSO2** | **PSO3** |
| **CO1** | **3** | **2** | **1** | **-** | **-** | **-** | **1** | **-** | **-** | **1** | **-** | **-** | **-** | **1** | **-** |
| **CO2** | **3** | **2** | **1** | **-** | **-** | **-** | **1** | **1** | **-** | **2** | **1** | **-** | **-** | **1** | **1** |
| **CO3** | **2** | **3** | **2** | **1** | **-** | **-** | **1** | **2** | **-** | **2** | **1** | **1** | **1** | **2** | **2** |
| **CO4** | **1** | **3** | **2** | **1** | **1** | **1** | **2** | **-** | **1** | **1** | **1** | **-** | **1** | **1** | **2** |
| **CO5** | **3** | **2** | **1** | **1** | **-** | **1** | **1** | **-** | **2** | **-** | **-** | **1** | **1** | **-** | **1** |

### High-3 Medium-2 Low-1

### UNIT I: Data Warehousing and Business Analysis 9 Hours

### Data Warehousing and Business Analysis: - Data warehousing Components –Building a Data warehouse –Data Warehouse Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata – reporting – Query tools and Applications – Online Analytical Processing (OLAP) – OLAP and Multidimensional Data Analysis.

### UNIT II: Data Mining 9 Hours

### Data Mining: - Data Mining Functionalities – Data Preprocessing – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization and Concept Hierarchy Generation- Architecture Of A Typical Data Mining Systems- Classification Of Data Mining Systems.

### UNIT III: Classification and Prediction 9 Hours

### Classification and Prediction: - Issues Regarding Classification and Prediction – Classification by Decision Tree Introduction – Bayesian Classification – Rule Based Classification – Classification by Back propagation Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction – Accuracy and Error Measures – Evaluating the Accuracy of a Classifier or Predictor – Ensemble Methods – Model Section.

### UNIT IV: Cluster Analysis 9 Hours

### Cluster Analysis: - Types of Data in Cluster Analysis – A Categorization of Major Clustering Methods – Partitioning Methods – Hierarchical methods – Density-Based Methods – Grid-Based Methods – Model-Based Clustering Methods – Clustering High-Dimensional Data – Constraint-Based Cluster Analysis – Outlier Analysis.

### UNIT V: Mining Object 9 Hours

### Mining Object, Spatial, Multimedia, Text and Web Data: Multidimensional Analysis and Descriptive Mining of Complex Data Objects – Spatial Data Mining – Multimedia Data Mining – Text Mining – Mining the World Wide Web.

**Total: 45 Hours**

**Reference (s)**

1. Alex Berson and Stephen J. Smith “Data Warehousing, Data Mining & OLAP”, Tata McGraw – Hill Edition, Tenth Reprint 2007.
2. K.P. Soman, Shyam Diwakar and V. Ajay “Insight into Data mining Theory and Practice”, Easter Economy Edition, Prentice Hall of India, 2006.
3. G. K. Gupta “Introduction to Data Mining with Case Studies”, Easter Economy Edition, Prentice Hall of India, 2006.

**List of e-Learning Resources:**

1. https://nptel.ac.in/
2. <https://www.coursera.org/>

**Subject Tr. Academic Coordinator HoD Sr. Faculty Nominated by DOAA**