

|                           |   |
|---------------------------|---|
| <b>Branch: B.Sc.(IT)</b>  | <b>Semester-IV</b>                      |
| <b>Subject Code: 4103</b> | <b>Lecture: 04</b><br><b>Credit: 04</b> |
| <b>Course Opted</b>       | <b>Core Course - 13</b>                 |
| <b>Subject Title</b>      | <b>DATA WAREHOUSING AND DATA MINING</b> |

**Course Objectives:**

- Understand the necessity of Data Warehousing and its continuous growth.
- Understand Planning and Management of Data Warehouse.
- Understand issues in various Architectural types of Data warehouse.
- Understand the application of various models of Data Warehouse.
- Understand the web-enabled data warehouse and role of data mining

**Course Outcomes:**

On completion of the course the student will be able to

- Decide the type of Data warehouse to build.
- Perform Requirement gathering and Design suitable architecture for Data warehouse project.
- Design and prepare data for Data warehouse using ETL tools
- Build web-enabled data warehouse
- Analyze and Apply Data Mining techniques on real life applications
- Demonstrate phases in data warehouse development life cycle with Data warehouse project.

| <b>Modules</b>  | <b>Sr. No.</b> | <b>Topic and Details</b>   | <b>No of Lectures Assigned</b> | <b>Marks Weightage %</b> |
|-----------------|----------------|--|--------------------------------|--------------------------|
| <b>UNIT - I</b> | 1              | <b>Introduction to Data Mining:</b><br>Basic Data Mining Tasks, DM versus Knowledge Discovery in Databases, Data Mining Issues, Data Mining Metrics, Social Implications of Data Mining, Overview of Applications of Data Mining | 4                              | 8                        |
|                 | 2              | <b>Data ware housing building blocks:</b><br>Data ware house and Data Marts, Architecture of DW, Components, Metadata, OLAP and Data Cubes   | 4                              | 8                        |
|                 | 3              | <b>Data Warehousing Concept:</b><br>Dimensional Data Modeling-star, snowflake schemas, Data Preprocessing – Need, Data Cleaning, Data Integration & Transformation, Data Reduction, Machine Learning, Pattern Matching           | 4                              | 8                        |
|                 |                | <b>Data Mining Techniques:</b><br>Frequent item-sets and Association rule mining: Apriori algorithm, Use of sampling for frequent item-set, FP tree algorithm, Graph Mining: Frequent sub-graph mining, Tree mining              | 4                              | 10                       |

|              |    |  |           |            |
|--------------|----|--|-----------|------------|
|              |    | Sequence Mining  |           |            |
| UNIT - II    | 5  | <b>Classification &amp; Prediction</b><br>Decision tree learning: Construction, performance, attribute selection, Issues: Over-fitting, tree pruning methods, missing values, continuous classes, Classification and Regression Trees (CART)<br>Bayesian Classification: Bayes Theorem, Naïve Bayes classifier, Bayesian Networks, Inference, - Parameter and structure learning, -Linear classifiers, -Least squares, logistic, perceptron and SVM classifiers, Prediction, -Linear regression, - Non-linear regression | 10        | 24         |
| UNIT - III   | 6  | <b>Accuracy Measures:</b><br>Precision, recall, F-measure, confusion matrix, cross-validation, bootstrap   | 4         | 6          |
|              | 7  | <b>Software for data mining and applications of data mining :</b><br>R, Weka, Sample applications of data mining   | 6         | 8          |
|              | 8  | <b>Clustering:</b><br>k-means, Expectation Maximization (EM) algorithm, Hierarchical clustering, Correlation clustering  | 4         | 8          |
| UNIT - IV    | 9  | <b>Brief overview of advanced techniques:</b><br>Active learning, Reinforcement learning, Text mining, Graphical models, Web Mining  | 4         | 8          |
|              | 10 | <b>Case Studies:</b><br>Discuss test cases for data warehouse applications, Discovering web access patterns and trends by Data Mining Technology on Web Logs, Discovering web access patterns and trends by applying OLAP, Discuss design of data warehouse/ Data Mart for suitable system   | 6         | 12         |
| <b>TOTAL</b> |    |  | <b>50</b> | <b>100</b> |

**Text Book:**

1. Data Mining and Data Warehousing: Principles and Practical Techniques , by Parteek Bhatia.(2019)

**Reference Books:**

1. Data Mining: Concepts and Techniques, Jiawei Han, Micheline Kamber, Jian Pei.(2012)
2. Data Mining – Introductory and Advanced Topics, Pearson Education, Margaret H. Dunham, S. Sridhar(2020).
3. Christopher M. Bishop, —Pattern Recognition and Machine Learning, Springer

4. Ian H.Witten, Eibe Frank Data Mining: Practical Machine Learning Tools and Techniques, Elsevier/(Morgan Kaufman), ISBN:9789380501864(2011)
- 5.Hands on Programming with R,Oreilly,Garrett Golemund by Hadley Wickham.(2014)

**Web References:**

1. <https://www.cs.waikato.ac.nz/ml/weka>
2. <https://nptel.ac.in/courses/111/104/111104120/>