Faculty of Science and Technology
R.T.M Nagpur University, Nagpur
Syllabus for B.Tech. Sixth Semester CT
Data Warehousing and Mining (Theory)

Total Credits: 03	Subject Code: BTCT602T
Teaching Scheme : **Lectures: 03 Hours/Week** **Tutorials: 00 Hours/Week** **Practical: 00 Hours/Week**	**Examination Scheme :** **Duration of University Exam : 03 Hrs.** **College Assessment : 30 Marks** **University Assessment:70 Marks**

Course Objectives:

To make students

1. To understand the basic concepts of Data Warehouse and Data Mining techniques.
2. Capable to create a data warehouse and to process raw data .
3. Able to apply basic classification, clustering on a set of data.
4. Able to identify frequent data items and to apply association rule on a set of data.
5. To learn recent trends of data mining such as web mining.

Course Outcomes:

After completion of the course, students will be able to -

1. Understand the data warehousing components and design a data warehouse for any organization.
2. Learn data mining concepts and working.
3. Explore functionality of the various data mining techniques.
4. Discuss the data-mining tasks like classification, clustering, association mining and extract knowledge using data mining techniques.
5. Apply data mining techniques in trending domain such as web mining and Solve real-world problems in business and scientific information using data mining.

UNIT I **(09 Hrs)**

Introduction: Characteristics, Operational database systems and data warehouse (OLTP & OLAP), Multidimensional data models, Data warehouse architecture, OLAP Operations, Design and construction of data warehouses.

UNIT II **(06 Hrs)**

Fundamentals of data mining:Data mining functionalities, Classification of data mining systems, Data mining task primitives, Major issues and challenges in data mining, Data preprocessing- need for processing, data cleaning, integration, transformation, data reduction, data mining application areas.

UNIT III **(09 Hrs)**

Classification: Introduction, Decision tree, Building decision tree- tree induction algorithm, Split algorithm based on information theory, Split algorithm based on gini index, Decision tree rules, Naive based methods.
Clustering: Cluster analysis, Desired features, Types of data in cluster analysis, Computing distance. Categorizations of major clustering methods – Partitioning methods (K-means, EM), Hierarchical methods (agglomerative, divisive).

UNIT IV **(06 Hrs)**

Mining frequent patterns and Association Rules: Market basket analysis, Frequent item sets and association rules, Apriori algorithm, FP growth algorithm, Improving efficiency of Apriori and FP growth algorithms.

UNIT V **(06 Hrs)**

Web Data Mining: Introduction, Graph properties of web, Web content mining, Web structure mining, Web usage mining, Text mining, Visual web data mining, Temporal and Spatial data mining.

TEXT BOOK:

1. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Second Edition, Elsevier, Reprinted 2008.
2. A. K. Pujari, "Data Mining Techniques", Second Edition, University press, 2013.
3. Jason Bell, "Machine Learning for Big Data: Hands-on for Developers and Technical Professionals, Wiley India Publications, 2013.

Data Warehousing and Mining (Practical)

Total Credits: 01	**Subject Code: BTCT602P**
Teaching Scheme: **Lectures: 00 Hours/Week** **Tutorials: 00 Hours/Week** **Practical: 02 Hours/Week**	**Examination Scheme:** **College Assessment: 25 Marks** **University Assessment:25 Marks**

Minimum ten experiments should be conducted based on the theory syllabus.