MCA-301 **Data Mining Using R** L=3, T=1

Course Objectives: The objective of this course is to provide the in-depth coverage of data mining and integration aspects along with its implementation in R programming language.

UNIT - I

Data Warehouse: A Brief History, Characteristics, Architecture for a Data Warehouse. Data Mining: Introduction, Motivation, Importance, Knowledge Discovery Process, Data Mining Functionalities, Interesting Patterns, Classification of Data Mining Systems, Major issues; Data Pre-processing: Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization, Outliers.

UNIT - II

Data Mining Techniques: Clustering-Requirement for Cluster Analysis, Clustering Methods- Partitioning Methods, Hierarchical Methods, Decision Tree-Decision Tree Induction, Attribute Selection Measures, Tree Pruning. Association Rule Mining-Market Basket Analysis, Frequent Itemset Mining using Apriori Algorithm, Improving the Efficiency of Apriori. Concept of Nearest Neighborhood and Neural Networks.

UNIT - III

Data Integration: Architecture of Data Integration, Describing Data Sources: Overview and Desiderate, Schema Mapping Language, Access Pattern Limitations, String Matching: Similarity Measures, Scaling Up String Matching, Schema Matching and Mapping: Problem Definition, Challenges, Matching and Mapping Systems, Data Matching: Rule- Based Matching, Learning- Based Matching, Matching by Clustering.

UNIT - IV

R Programming: Advantages of R over other Programming Languages, Working with Directories and Data Types in R, Control Statements, Loops, Data Manipulation and integration in R, Exploring Data in R: Data Frames, R Functions for Data in Data Frame, Loading Data Frames, Decision Tree packages in R, Issues in Decision Tree Learning, Hierarchical and K-means Clustering functions in R, Mining Algorithm interfaces in R.

Text Books:

1. J Hanes, M. Kamber, Data Mining Concepts and Techniques, Elsevier India.
2. A.Doan, A. Halevy, Z. Ives, Principles of Data Integration, Morgan Kaufmann Publishers.

3. S. Acharya, Data Analytics Using R, McGraw Hill Education (India) Private Limited.

Reference Books:

1. G.S. Linoff, M.J.A. Berry, Data Mining Techniques, Wiley India Pvt. Ltd.
2. Berson, S.J. Smith, Data Warehousing, Data Mining & OLAP, Tata McGraw-Hill.
3. J.Horbulyk, Data Integration Best Practices.
4. Jared P. Lander, R For Everyone, Pearson India Education Services Pvt. Ltd

Course Outcomes:

By the end of the Course, Student will be able to:

CO1:understand the fundamental concepts of data warehousing and data mining;

CO2:acquire skills to implement data mining techniques;

CO3:learn schema matching, mapping and integration strategies;

CO4:implement data mining techniques in R to meet the market job requirements.

NOTE: In Each theory paper. Nine questions are to be set. Two questions are to be set from each Unit and Candidate is required to attempt one question from each unit. Question number nine is Compulsory, which will be of short answer type with 5-10 parts, out of the entire syllabus. In all five questions are to be attempted.