4MCACC3: DATA WAREHOUSING AND KNOWLEDGE MINING

Total No. of Hours: 52 **Hours/Week: 04**

Course Objective: To introduce the concept of Data Mining techniques and its applications.

Course Outcome: Students will be able to

CO1: Understand the concepts of data warehouse and data mining

CO2: Use data pre - processing techniques to build data warehouse

CO3: Analyse transaction databases for association rules.

CO4: Use classification methods and prediction techniques on transaction databases.

CO5: Understand various clustering techniques for categorizing data.

CO6: Understand methods for outlier analysis.

Unit I	**Data Warehousing and Online Analytical Processing**: Basic concepts, Data warehouse Modelling, Data cube and OLAP, Date Warehouse Design and Usage, Data warehouse implementation. **Data Pre-processing:** Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization.	12 hrs
Unit II	**Data Mining**: Introduction, Kinds of Data, Patterns and Technologies, Architecture of Data Mining Systems, Applications, Primitives and Issues in Data Mining. **Exploring the Data**: Data Objects and Attributes, Data Quality, Statistical Descriptions of Data, Measuring Data Similarity and Dissimilarity, Data Visualization.	8 hrs
Unit III	**Mining Frequent Patterns Associations and correlations**: Basic concepts, Frequent Itemset Mining Methods, Patterns evaluation Methods. **Advanced Pattern Mining**: Pattern Mining: A Road Map, Pattern Mining in Multilevel, Multidimensional Space, Constraint Based Frequent Pattern Mining, Mining High-Dimensional Data and Colossal Patterns, Mining Compressed or Approximate Patterns, Pattern Exploration and Application.	12 hrs
Unit IV	**Classification**: Basic Concepts, Decision tree induction, Bayes classification Methods, Bayesian Belief Networks, Rule Based Classification, Lazy Learners, Model Evaluation and Selection. **Clustering**: Clustering Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid Based Methods, Evaluation of Clustering.	12 hrs
Unit V	**Data Mining Trends and Research Frontiers:** Mining complex Data types, Other Methodologies of Data Mining, Data Mining Applications, Data Mining and Society, Data Mining Trends. **Application:** Implementation using Data Mining tool.	8 hrs

REFERENCE BOOKS

[1] Jaiwei Han and Micheline Kamber, "*Data Mining: Concepts and Techniques*", Morgan Kaufman Publishers, Third Edition, San Francisco, USA 2002.

[2] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining", Addison-Wesley, 2006.

[3] Arun K Pujari, *"Data Mining Techniques", University* Press 2nd Edition, 2009

[4] Alex Berson and Stephen J. Smith, "Data *Warehousing, Data Mining & OLAP*", Tata McGraw – Hill Edition, Tenth Reprint 2007.

[5] Claudia Imhoff, Nicholas & et al "*Mastering Data Warehouse Design*", J. Wiley.