II Year MCA II Semester	L	T	P	To	C
	3	1	-	4	4

MC204 DATA WAREHOUSING AND MINING

Objectives of the Course:

The course aims to help students

- *To understand and implement classical algorithms in data mining and data warehousing.*
- *To assess the strengths and weaknesses of the algorithms.*
- *To identify the application area of algorithms, and to apply them.*

UNIT - I **(15 Hrs)**

Data Warehouse: Introduction, A Multi-dimensional data model, Data Warehouse Architecture, Data Warehouse Implementation, From Data Warehouse to Data Mining.

Data Mining – Introduction, Data Mining, Kinds of Data, Data Mining Functionalities, Classification of Data Mining Systems, Major issues in Data Mining.

UNIT - II **(12 Hrs)**

Data Preprocessing: Data cleaning, Data Integration & Transformation, Data Reduction, Discritization & Concept Hierarchy Generation, Data Mining Primitives.

Mining Association rules in large databases – Association rule mining, mining single-dimensional Boolean Association rules from Transactional Databases, Mining Multi-dimensional Association rules from relational databases & Data Warehouses.

UNIT - III **(10 Hrs)**

Concept Description: Introduction, Data Generalization and Summarization-Based Characterization, Analytical Characterization, Mining Class Comparisons, Mining Descriptive Statistical Measures in Large Databases.

UNIT - IV **(12 Hrs)**

Classification & Prediction: Introduction, Classification by Decision tree induction, Bayesian Classification, , Classification by Back propagation, Other Classification Methods, Prediction, Classifier accuracy.

Mining Complex Type of Data – Multidimensional Analysis and Descriptive Mining of Complex Data Objects, Mining Spatial Databases, Mining Multimedia Databases, Mining Text Databases, Mining the World Wide Web.

UNIT - V **(10 Hrs)**

Cluster Analysis: Introduction, Types of data in Cluster analysis, A categorization of major clustering methods, partitioning methods, Hierarchical methods, Density-Based Methods: DBSCAN, Grid-based Method: STING; Model-based Clustering Method: Statistical approach, Outlier analysis.

Text Books:

1. Data Mining Concepts & Techniques – Jiawei Han Micheline Kamber – Morgan Kaufmann Publishers.
2. Paulraj Ponnaiah, “Data Warehousing Fundamentals”, Wiley Publishers, 2001.

Reference Books:

1. Usama M.Fayyad, Gregory Piatetsky Shapiro, Padhrai Smyth, RamasamyUthurusamy, “Advances in Knowledge Discover and Data Mining”, The M.I.T. Press, 1996.
2. Ralph Kimball, Margy Ross, “The Data Warehouse Toolkit”, John Wiley and Sons Inc., 2002.
3. Alex Berson, Stephen Smith, Kurt Thearling, “Building Data Mining Applications for CRM”, Tata McGraw Hill, 2000.
4. Margaret Dunham, “Data Mining: Introductory and Advanced Topics”, Prentice Hall, 2002.