

(CS602) DATA MINING AND DATA WAREHOUSING

Objective of the course :

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 apply them

UNIT - I

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

Data Preprocessing : Data cleaning, Data Integration & Transformation, Data Reduction, Discretization & 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

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

UNIT - IV

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

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 BOOK :

1. Jiawei Han Micheline Kamber – “Data Mining Concepts & Techniques”, 1st ed., Morgan Kaufmann Publishers, 2007.

REFERENCE BOOKS :

1. Usama M.Fayyad, Gregory Piatetsky Shapiro, Padhraí Smyth, Ramasamy Uthurusamy, “Advances in Knowledge Discover and Data Mining”, 1st ed., The M.I.T. Press, 1996.
2. Ralph Kimball, Margy Ross, “The Data Warehouse Toolkit”, 1st ed., John Wiley and Sons Inc., 2002.
3. Alex Berson, Stephen Smith, Kurt Thearling, “Building Data Mining Applications for CRM”, 1st ed., Tata McGraw Hill, 2000.
4. Margaret Dunham, “Data Mining: Introductory and Advanced Topics”, 1st ed., Prentice Hall, 2002.
5. Paulraj Ponnaiah, “Data Warehousing Fundamentals”, 1st ed., Wiley Publishers, 2001.