Data Warehousing and Mining
BIT353CO

Year:IV   Semester:II

Teaching Schedule Hours/Week			Examination Scheme				Total
Theory	Tutorial	Practical	Internal		Final		
			Theory	Practical	Theory	Practical	150
3	1	2	20	50	80	-	

Course Objectives:

This course aims at introducing advance aspects of data warehousing and data mining, encompassing the principles, research results and commercial application of the current technologies. It also provides knowledge to introduce students to the basic concepts and techniques of data mining, using recent data mining software for solving practical problems.

Course Contents:

Unit 1   [4 Hrs]
Introduction to data mining: Motivation, importance, definition of data mining, kinds of data mining, kinds of patterns, data mining technologies, kinds of applications targeted, major issues in data mining; Introduction to Data warehousing: Importance, uses and applications

Unit 2   [9 Hrs]
Data Warehouse and OLAP Technology, Data Warehouse Architecture, Stepsfor the Design and Construction of Data Warehouses, A Three-Tier Data WarehouseArchitecture, OLAP, OLAP queries, metadata repository, Data Preprocessing – Data Integration and Transformation, Data Reduction, Data Mining Primitives: What Defines a Data Mining Task? Task-Relevant Data, The Kind of Knowledge to be Mined, KDD.

Unit 3   [9 Hrs]
Mining Association Rules in Large Databases, Association Rule Mining, Market Basket Analysis: Mining A Road Map, The Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation, Generating Association Rules from Frequent Itemsets, Improving the Efficiently of Apriori, Mining Frequent Itemsets without Candidate Generation, Multilevel Association Rules,

Approaches to Mining Multilevel Association Rules Mining

Unit 4   [6 Hrs]

Multidimensional Association Rules for Relational Database and Data Warehouses, Multidimensional Association Rules, Mining Quantitative Association Rules, Mining Distance-Based Association Rules, From Association Mining to Correlation Analysis

Unit 5   [9 Hrs]

What is Classification? What Is Prediction? Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Bayes Theorem, Naïve Bayesian Classification, Classification by Backpropagation, A Multilayer Feed-Forward Neural Network, defining a Network Topology, Classification Based of Concepts from Association Rule Mining, Other Classification Methods, k-Nearest Neighbor Classifiers, Genetic Algorithms, Rough Set Approach, Fuzzy Set approaches.

Unit 6   [8Hrs]

What Is Cluster Analysis? Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Classical Partitioning Methods: k-Meansand k-Medoids, Partitioning Methods in Large Databases: From k-Medoids to CLARANS, Hierarchical Methods.

Laboratory works:

The student must do the project work using data mining and data warehousing concept. Topics should be given by the course instructor and at the end of the semester student should present their project work.

Reference Books:

1. Morgan Kaufmann J. Han, M Kamber, "Data Mining Concepts and Techniques, Second edition
2. Sam Anahory, Dennis Murray, "Data Warehousing in the Real World", Pearson Education
3. P. and D. Zatinge, "Data Mining", Adriaans, Addison Wesley, 1996
4. Kimball, R., "The Data Warehouse Toolkit", Wiley 1996
5. W. H. Inmon, "Building The Data Warehouse", 3rd Edition, Wiley, 2003
6. Margaret H. Dunham, "Data Mining: Introductory and Advance Topics", Pearson Education 2004