

University of Mumbai			
Class: T.E.	Branch: Computer Engineering	Semester: VI	
Subject: DATA WAREHOUSING AND MINING (Abbreviated as DWM)			
Periods per Week (each 60 min)	Lecture	04	
	Practical	02	
	Tutorial	--	
		Hours	Marks
Evaluation System	Theory	03	100
	Practical and Oral	--	25
	Oral	---	--
	Term Work	---	25
	Total	03	150

Objectives: The data warehousing part of module aims to give students a good overview of the ideas and techniques which are behind recent development in the data warehousing and online analytical processing (OLAP) fields, in terms of data models, query language, conceptual design methodologies and storage techniques. Data mining part of the model aims to motivate, define and characterize data mining &S process; to motivate, define and characterize data mining applications.

Pre-requisites: DBMS

Module	Contents	Hours
Data Warehousing		
1	Overview and Concepts: Need for data warehousing, The building blocks of a Data warehouse.	04
2	Architecture and Infrastructure: Data Warehouse Architecture, Infrastructure and Metadata Management	04
3	Principles Of Dimension Modeling: Introduction to Dimensional Modeling, Advanced Concepts	04
4	Extract Transform Load Cycle: ETL overview, Extraction, Loading, Transformation techniques.	04
5	Information Access and Delivery: Matching information to classes of users, OLAP – the need, Design of the OLAP database, OLAP operations: slice, dice, rollup, drill-down etc. OLAP implementations.	04
6	Implementation And Maintenance: Physical design process, Aggregates and Indexing. Data Warehouse Deployment	04
Data Mining		
7	Introduction: Basics of data mining, related concepts, Data mining techniques. The KDD process	04

8	Concept Description: Class Characterization and comparison, Attribute relevance analysis, Attribute oriented Induction, Mining descriptive statistical measures in large databases.	04
9	Classification Algorithms: What is Classification? Supervised Learning, Classifier Accuracy, Decision Tree and Naïve Bayes Classifier.	04
10	Clustering: What is clustering? Types of data, Partitioning Methods (K-Means, K-Medoids) Hierarchical Methods(Agglomerative , Divisive)	04
11	Association rules: Motivation For Association Rule mining, Market Basket Analysis, Apriori Algorithm, FP tree Algorithm, Iceberg Queries. Advanced Association Rules (just concepts)	04
12	Web Mining: Web Content Mining, Web Structure Mining, Web Usage mining	04

BOOKS

Text Books:

- 1) Ralph Kimball, "The Data Warehouse Lifecycle toolkit', 2nd edition, Wiley India.
- 2) Han, Kamber, "Data Mining Concepts and Techniques", 2nd edition ,Elsevier
- 3) Reema Theraja "Data warehousing", Oxford University Press.
- 4) "Introduction to Data Mining", 1/e Pang-Ning Tan, Vipin Kumar, Michael Steinbach Pearson Education
- 5) M.H. Dunham, "Data Mining Introductory and Advanced Topics", Pearson Education.

Reference Books :

- 1) Paulraj Ponniah, "Data Warehousing Fundamentals", Wiley Student edition.
- 2) "Data mining For Business intelligence" Galit Shmueli, Nitin Patel, Peter Bruce; Wiley Student Edition.
- 3) "Data Warehousing, Data Mining & OLAP" Alex berson & Stephen J Smith, Tat McGraw Hill.
- 4) "Data Mining with SQL Server 2008" Jamie McLennan & others, Wiley Indian Edition.
- 5) "Mastering Data Mining", M Berry and G. Linoff, Wiley Student Edition.
- 6) R. Kimball, "The Data Warehouse Toolkit', John Wiley.

TERM-WORK

Term work should consist of at least of the following:

1. One case study given to a group of 3 /4 students, who will start form dimensional modeling and go upto generating OLAP reports..
2. Programming the data mining algorithms (classification, clustering and Association mining) in Java on example data sets. (Can compare with tools like WEKA).
3. Study of some BI tool like SQL SERVER or ORACLE etc.

PRACTICAL & ORAL EXAMINATION

Practical and Oral examination is based on the entire syllabus and may not be restricted to the practical carried out in the practical examination