EASWARI ENGINEERING COLLEGE, CHENNAI-600 089
DEPARTMENT OF INFORMATION TECHNOLOGY
LESSON PLAN

SUBJECT CODE : IF7203

SUBJECT TITLE : DATA WAREHOUSING AND DATA MINING

HOURS DISTRIBUTION : (L T P C 3 0 0 3)

COURSE/ BRANCH : M.E/ Software Engineering

SEMESTER : II

ACADEMIC YEAR : 2014 - 2015

FACULTY NAME : K.VALARMATHI

OBJECTIVE OF COURSE :

Data Mining studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. It is currently regarded as the key element of a more general process called *Knowledge Discovery* that deals with extracting useful knowledge from raw data. The course will cover all these issues and will illustrate the whole process by examples of practical applications. The students will use recent Data Mining software.

OUTCOME OF COURSE

At the end of the course student are able to:

1. Learn the concepts of database technology evolutionary path which has led to the need for data mining and its applications

2. Discover interesting patterns from large amounts of data to analyze and extract patterns to solve problems, make predictions of outcomes.

3. Evaluate systematically supervised and unsupervised models and algorithms with respect to their accuracy. Develop practical work of DM techniques and design hypotheses based on the analysis to conceptualize a DM solution to a practical problem.

4 Evaluate and select appropriate data-mining algorithms and apply, and interpret and report the output appropriately.Design and implement of a data-mining application using sample, realistic data sets and modern tools.

Format No. :LP-01

Issue No. : 01

Issue Date : 28.1.2012

EASWARI ENGINEERING COLLEGE

DEPARTMENT OF INFORMATION TECHNOLOGY

LESSON PLAN

Subject code: IF7203 **Degree/Branch: M.E** (Software Engg)

Subject Name: DATA WAREHOUSING AND DATA MINING

Year/Sem: I / II

Faculty Name: Mrs.K.Valarmathi

Lecture:45

Tutorial:0

Practical:0

Total:45 Hours

Sl.No	TOPIC	No.of Periods	Reference Books	Page No
	UNIT I			
	DATA WAREHOUSE			
Objective: To expose the students to the concepts of Data warehousing Architecture and Implementation				
1	Data Warehousing	1	R1	105-108
2	Operational Database Systems vs. Data Warehouses	1	R1	108-110
3	Multidimensional Data Model	1	R1	110-114
4	Schemas for Multidimensional Databases	1	R1	114-117

5	OLAP Operations	1	R1	123-126
6	Data WarehouseArchitecture	1	R1	127-137
7	Indexing	1	R1	141-144
8	OLAP queries & Tools.	1	R1	144-146
	Total	**8**		

Outcome: At the end of this module,the students will be able to Store voluminous data for online processing.

UNIT II

DATA MINING & DATA PREPROCESSING

Objective: To Understand Data mining principles and techniques and Introduce the DataMining as a cutting edge of business inteligence

9	Introduction to KDD proces –	1	R1	5-9
10	Knowledge Discovery from Database	1	R1	9-21
11	Need for Data Pre-processing-why preprocessing need	1	R1	47-51
12	Descriptive Data summarization	1	R1	51-61
13	Data Cleaning	1	R1	61-67
14	Data Integration and Transformation	1	R1	67-72
15	Data Reduction	1	R1	72-86
16	DataDiscretization	1	R1	87-94
17	Concept Hierarchy Generation.	1	R1	94-96
	Total	**9**		

Outcome: At the end of this module, the students will be able to know how to preprocess the data for mining applications.

UNIT III

ASSOCIATION RULE MINING

Objective: To learn to use association rule mining for handling large data.

18	Introduction	1	R1	21
19	Data Mining Functionalites	1	R1	21-27
20	Association Rule Mining - Aprior Algorithm,Generating association rules - Improving efficiency of Aprior	1	R1	234-242
21	Mining Frequent Itemsets without Candidate Generation	1	R1	242-245
22	Mining Frequent Itemsets vertical data format,Mining closed frequent itemset	1	R1	245-250
23	Mining Various Kinds of Asociation Rules -Mining multilevel association rule	1	R1	250-254
24	Mining Various Kinds of Asociation Rules -Mining multidimensional Association rule	1	R1	254-259
25	Constraint-Based Asociation Mining	1	R1	265-272
	Total	**8**		
Outcome: At the completion of this unit, the students will be able to apply the association rules for mining the data.				
UNIT IV				
CLASSIFICATION & PREDICTION				
Objective: To understand the concept of classification for the retrieval purposes.				
26	Classifcation vs. PredictionClassifcation by Decision Tree Induction	1	R1	291-310
27	Bayesian Classifcation	1	R1	310-318
28	Rule Based Classifcation	1	R1	318-327
29	Classifcation by Back Propagation	1	R1	327-337
30	Support Vector Machines	1	R1	337-344
31	Associative Classifcation	1	R1	344-347

32	Lazy Learners	1	R1	347-351
33	Other Classifcation Methods	1	R1	351-354
34	Prediction – Accuracy and Error Measures	1	R1	354-363
35	Evaluating the Accuracy of a Classifer or Predictor ,Ensemble Methods , Model Section	1	R1	363-373
	Total	**10**		

Outcome:At the end of this unit,students will be able to Design and deploy appropriate classifcation techniques

UNIT V

CLUSTERING

Objective:

- To know the clustering techniques in details for better organization and retrieval of data
- To identify Business applications and Trends of Data mining

36	Cluster Analysis	1	R1	383-386
37	Types of Data in Cluster Analysis	1	R1	386-398
38	Categorization of Major Clustering Methods	1	R1	398-401
39	Partioning Methods	1	R1	401-408
40	Hierarchical methods	1	R1	408-418
41	Density-Based Methods	1	R1	418-424
42	Grid-Based Methods	1	R1	424-429
43	Model-Based Clustering Methods	1	R1	429-434
44	Clustering High-Dimensional Data –	1	R1	434-444
45	Constraint- Based Cluster Analysis – Outlier Analysis	1	R1	444-460
	Total	**10**		

Outcome: The students wil be able to, Cluster the high dimensional data ,Discover the knowledge imbibed in the high dimensional system , evolve Multidimensional Intelligent model from typical system and Evaluate various mining techniques on complex data objects

BEYOND THE SYLLABUS

1. Graph Mining
2. Current trends in Data Warehousing.

ASSIGNMENTS

1. Data Mining for Financial Data Analysis.
2. Social Network Analysis.
3. Examples of Commercial data mining System.

References

1. Jiawei Han and MichelineKamber, “Data Mining Concepts and Techniques” Second Editon,
Elsevier, Reprinted 2008.
2. K.P. Soman, ShyamDiwakar and V. Ajay, “Insight into Data mining Theory and Practice”,
Easter Economy Editon, Prentice Hal of India, 206.

3. G. K. Gupta, “Introduction to Data Mining with Case Studies”, Easter Economy Editon, Prentice Hal of India, 206.
4. BERSON, ALEX & SMITH, STEPHEN J, Data Warehousing, Data Mining, and OLAP, TMH
Pub. Co. Ltd, New Delhi, 2012
5. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, “Introduction to Data Mining”, Pearson
Education, 207

6. PRABHU Data Warehousing, PHI Learning Private Limited, New Delhi, 2012,

7. PONNIAH, PAULRAJ, Data Warehousing Fundamentals, John Wiley & Sons, New Delhi,
2011

Prepared By **Approved By**

Program Educational Outcomes

1. Graduates will be proficient in utilizing the fundamental knowledge of basic sciences and mathematics to the applications relevant to various streams of Engineering and Technology.

2. Graduates will possess core competencies necessary for application of knowledge of computers and telecommunications equipment to store, retrieve, transmit, manipulate and analyze data in the context of business enterprise.

3. Graduates will be capable of thinking logically, pursue lifelong learning and will have the capacity to understand technical issues related to computing systems and design optimal solutions.

4. Graduates will be able to develop hardware and software systems by understanding the importance of social, business and environmental needs in the human context.

5. Graduates will gain employment in organizations and establish themselves as professionals by applying their technical skills to solve real world problems and meet the diversified needs of industry, academia and research.

6. Graduates will be aware of professional ethics of the software industry and equip themselves with communication skills essential for working in community.

Program Outcomes

(a) Ability to apply knowledge of computing and mathematics appropriate to Information Technology

(b) Ability to analyze a problem, and identify computing requirements appropriate to its solution

(c) Ability to design, implement, and evaluate a system, process, component, or program to meet specific requirements

(d) Ability to interpret and synthesis data to provide valid conclusions

(e) Ability to function effectively as a team member to achieve a common goal

(f) Ability to understand professional, ethical and social issues and responsibilities

(g) Ability to communicate effectively with a diverse groups

(h) Ability to analyze the local and global impact of Information Technology on society

(i) Ability to recognize and engage in continuing professional development and life long learning

(j) Ability to use current techniques, skills, and tools necessary to accomplish projects related to Information Technology.

(k) Ability to understand the impact of the professional engineering solutions in societal and environmental contexts for sustainable development.

(l) Ability to understand engineering and management principles to manage projects in multidisciplinary environment.

UNITS	*Course outcome*	*PEO 1*	*PEO 2*	*PEO 3*	*PEO 4*	*PEO 5*	*PEO 6*	*PO a*	*PO b*	*PO c*	*PO d*	*PO e*	*PO f*	*PO g*	*PO h*	*PO i*	*POj*	*POk*	*POl*
DATA WAREHOUSE	At the end of this module, the students will be able to Store voluminous data for online processing.	w	s	w	M	s	W	W	S	M	S	W	W	W	M	W	W	W	W
DATA MINING & DATA PREPROCESSING	At the completion of this unit, the students will be able to apply the association rules for mining the data.	w	W	S	S	W	W	W	S	S	S	W	W	W	M	W	W	W	W
ASSOCIATION RULE MINING	At the completion of this unit, the students will be able to apply the association rules for mining the data.	w	s	W	S	W	W	W	S	S	S	W	W	W	W	W	W	W	W

CLASSIFICATION& PREDICTION	At the end of this unit,students will be able to Design and deploy appropriate classifcation techniques	w	s	w	M	s	s	W	S	S	S	M	M	W	W	W	W	W	W
CLUSTERING	The students wil be able to, Cluster the high dimensional data ,Discover the knowledge imbibed in the high dimensional system ,evolve Multidimensional Intelligent model from typical system and Evaluate various mining techniques on complex data objects	w	s	S	S	M	s	W	M	S	S	W	W	W	M	W	W	W	W

STRONG	S
MEDIUM	M
WEAK	W