

COURSE HANDOUT

B. Tech (CSE) – 6th Semester

Course Title : **Data Engineering**

Dated: 20-11-2017

Course Code : CSE 3412

Academic Year: 2017-18

Course Structure : 3-1-0-4

Course Coordinator: Mrs. N. Lakshmi Devi

Instructor(s) : Mr. K. Somesh, Mrs. N. Lakshmi Devi, Mr. K. Suresh Kumar

Pre-requisite : Mathematics, Statistics and Database Management Systems

Course Description:

Data that has relevance for managerial decisions is accumulating at an incredible rate due to a host of technological advances. Electronic data capture has become inexpensive and ubiquitous as a by-product of innovations such as the internet, e-commerce, electronic banking, point-of-sale devices, bar-code readers, and intelligent machines. Such data is often stored in data warehouses and data marts specifically intended for management decision support. Data mining is a rapidly growing field that is concerned with developing techniques to assist managers to make intelligent use of these repositories. A number of successful applications have been reported in areas such as credit rating, fraud detection, database marketing, customer relationship management, and stock market investments. The field of data Engineering has evolved from the disciplines of statistics and artificial intelligence.

This course will examine methods that have emerged from both fields and proven to be of value in recognizing patterns and making predictions from an applications perspective. The applications of Data Engineering provide an opportunity for hands-on experimentation with algorithms for data mining.

Scope and Objective:

Data Engineering is a multidisciplinary field, drawing work from area including database technology, artificial intelligence, machine learning, neural networks, and statistics. Data Mining is a process of discovering interesting knowledge from large amounts of data stored in databases and data warehouses.

The course content enables students to:

1. Understand fundamentals of data mining concepts.
2. Understand the concept of association rule mining and generate association rules.
3. Learn classification algorithms and classify the real data set.
4. Analyze clustering methods and understand mining complex data types.

Text Books:

1. Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson Education.
2. Data Mining: Concepts and Techniques, Jiawei Han & Micheline Kamber, Morgan Kaufmann, India.

Reference books:

1. Data Mining: Introductory and Advanced Topics, Margaret H. Dunham, Pearson Education.
2. Data Mining Techniques, Arun K. Pujari, University Press.

SYLLABUS:

UNIT I:

15 Hrs

Introduction: What is data warehouse, Data warehouse architecture, What is data mining, Motivating challenges, the origins of data mining, Data Mining Tasks.

Data: Types of data, Data quality, Data preprocessing, Measures of similarity and dissimilarity - Basics, Similarity and Dissimilarity between Simple Attributes, Similarities and Dissimilarities between Data objects, Examples of proximity measures, Issues in proximity Calculations, selection of the right proximity Measure.

UNIT II:

16 Hrs

Mining Association Rules in Large Databases: Association Rule Mining, Mining Single-Dimensional Boolean Association Rules from Transactional Databases, Mining Multilevel Association Rules from Transaction Databases, Mining Multidimensional Association Rules from data sources, From Association Mining to Correlation Analysis, Constraint-Based Association Mining.

UNIT III:

13 Hrs

Classification: Preliminaries, General approach to solving a classification problem, Decision tree induction, Rule-based classifier, Nearest-Neighbor classifiers, Bayesian classifiers, Artificial Neural Network.

UNIT IV:

16 Hrs

Cluster Analysis: Overview, K-means, Agglomerative Hierarchical Clustering, DBSCAN, Prototype-based clustering, Cluster Evaluation.

Mining Complex Types 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.

Course Outcome: At the end of the course students will be able to:

1. Apply relevant preprocessing techniques on different data sets for mining.
2. Implement Association rule mining concept and generate association rules.
3. Apply classification/clustering techniques on different types of data and analyze patterns.
4. Suggest appropriate data mining techniques to mine different types of data.

Course Plan:

No. Lecture	Learning Objectives	Topic(s) to be Covered	Chapter in the Textbook/ Reference
Unit-1: Introduction & Data			
1.	To learn the Concept of Data mining	Introduction to Data Mining	1-1.1,2-1
2.	To know the basic Terminology of Data Warehouse Architecture	What is Data Warehouse, Data Warehouse Architecture	2-3.1,3.3
3.	To study the basic Terminology of Data mining, Major Challenges and Issues in Data mining	What is Data mining, Motivating Challenges	1-1.1 & 1.2, 2-1

4.	TUTORIAL – 1		
5.	To know the Origins of Data mining and its Tasks	The origins of Data mining, Data Mining Tasks	1-1.3,1.4
6.	To learn the various Data Types and their Quality	Types of Data & Data Quality	1-2.1, 2.2
7.	To learn needs of Preprocessing Data	Data Preprocessing	1-2.3,2-2
8.	TUTORIAL – 2		
9.	To understand the Concept of Similarity and Dissimilarity	Measures of Similarity and Dissimilarity - Basics	1-2.4,2-2
10.	To implement the concept of Similarity and Dissimilarity in Simple Attributes	Similarity and Dissimilarity between Simple Attributes	1-2.4,2-2
11.	To Identify the concept of Similarity and Dissimilarity in Data Objects	Similarities and Dissimilarities between Data objects	1-2.4,2-2
12.	TUTORIAL – 3		
13.	To learn and implement the Proximity Measures	Examples of Proximity Measures, Issues in Proximity Calculations	1-2.4,2-2
14.	To identify the Proximity Measures for specific data	Selection of the Right Proximity Measure	1-2.4,2-2
15.	TUTORIAL – 4		
Unit-2: Mining Association Rules in Large Databases			
16.	To find Frequent Patterns in Data	Association Rule Mining-Concept	1-6.1, 2-5.1
17.	To understand and implement the Single-dimensional Boolean Association Rules from Transactional Database	Mining Single-Dimensional Boolean Association Rules from Transactional Databases	1-6.2, 2-5.2
18.	To understand and implement the Single-dimensional Boolean Association Rules from Transactional Database	Mining Single-Dimensional Boolean Association Rules from Transactional Databases-Examples	1-6.2, 2-5.2
19.	TUTORIAL – 5		
20.	To understand the Multilevel Association Rules from Transactional Database	Mining Multilevel Association Rules from Transaction Databases	2-5.3
21.	To implement the Multilevel Association Rules from Transactional Database	Mining Multilevel Association Rules from Transaction Databases-Examples	1-6.6, 2-5.3
22.	To understand the Multidimensional Association rules from Transactional Database	Mining Multidimensional Association Rules from data sources	1-6.6, 2-5.3
23.	TUTORIAL – 6		
24.	To implement the Multidimensional Association rules from transactional database	Mining Multidimensional Association Rules from data sources-Examples	1-7.1, 2-5.3
25.	To understand the basic concept on Association Mining and Correlation Analysis	From Association Mining to Correlation Analysis	2-5.4
26.	To implement the Association Mining and Correlation Analysis	From Association Mining to Correlation Analysis-Examples	1-6.7 & 7.6, 2-5.4
27.	TUTORIAL – 7		

28.	To understand the Constraint-Based Association Mining	Constraint-Based Association Mining	2-5.5
29.	To implement the Constraint-Based Association Mining	Constraint-Based Association Mining-Examples	2-5.5
30.	To differentiate the different Association Rule Mining Concepts	Overview of different Association Rule Mining Concepts	1, 2-5
31.	TUTORIAL – 8		
Unit-3: Classification			
32.	To know the basic Concept of classification	Preliminaries, General approach to solving a Classification Problem	1-4.1& 4.2, 2-6.1
33.	To learn and implement Decision Tree Induction to Generate Classification Rules	Decision Tree Induction	1-4.3, 2-6.3
34.	To implement Decision Tree Induction to generate Classification Rules	Decision Tree Induction-Examples	1-4.3, 2-6.3
35.	TUTORIAL – 9		
36.	To learn and implement Rule-based Classifier to generate Classification Rules	Rule-based Classifier	1-5.1, 2-6.5
37.	To learn and implement Nearest-Neighbor Classifiers to generate Classification Rules	Nearest-Neighbor Classifiers	1-5.2, 2-6.9.1
38.	To learn and implement Rule-based & Nearest-Neighbor Classifiers to generate Classification Rules	Rule-based & Nearest-Neighbor Classifiers to generate Classification Rules- Examples	1-5.1, 2-6.5 & 1-5.2, 2-6.9.1
39.	TUTORIAL – 10		
40.	To learn Bayesian Classifiers to generate Classification Rules	Bayesian classifiers	1-5.3, 2-6.4
41.	To learn Artificial Neural Network to generate Classification Rules	Artificial Neural Network	1-5.4, 2-6.6
42.	To implement Bayesian and Artificial Neural Network classifiers to generate Classification Rules	Bayesian and Artificial Neural Network classifiers to generate Classification Rules - Examples	1-5.3, 2-6.4 & 1-5.4, 2-6.6
43.	TUTORIAL – 11		
44.	To differentiate the different classifiers to generate Classification Rules Concepts	Overview of different Classifiers Concepts	1-5,2-6
Unit – 4: Cluster Analysis & Mining Complex Types of Data			
45.	To know the process of Partitioning a Set of Data Objects into subsets	Overview of Cluster Analysis	1-8.1, 2-7.1&7.2
46.	To learn and implement K-means Algorithm for Cluster Analysis	K-means Algorithm	1-8.2, 2-7.4
47.	To learn and implement Agglomerative Hierarchical Clustering Algorithm for Cluster Analysis	Agglomerative Hierarchical Clustering Algorithm	1-8.3, 2-7.5.1
48.	TUTORIAL – 12		

49.	To learn and implement DBSCAN Algorithm for Cluster Analysis	DBSCAN Algorithm	1-8.4, 2-7.6.1
50.	To learn and implement Prototype-based clustering Algorithm for Cluster Analysis	Prototype-based clustering Algorithm	1-9.2
51.	To Evaluate the Clusters in Cluster Analysis	Cluster Evaluation	1-8.5, 2-7.8
52.	TUTORIAL – 13		
53.	To learn about Multidimensional Analysis	Multidimensional Analysis	1-3.4, 2-10.1
54.	To learn about Complex Data Objects and its mining	Descriptive Mining of Complex Data Objects	2-10.1
55.	To learn Spatial Databases Mining	Mining Spatial Databases	2-10.2
56.	TUTORIAL – 14		
57.	To learn Multimedia Databases Mining	Mining Multimedia Databases	2-10.3
58.	To learn Text Databases Mining	Mining Text Databases	2-10.4
59.	To learn the World Wide Web Databases Mining	Mining the World Wide Web	2-10.5
60.	TUTORIAL – 15		

Evaluation Scheme:

Component	Duration (minutes)	Marks	% of weightage	Date & Time	Venue
Sessional Test – 1	90	20	20%	18.01.2016 to 23.01.2016	Block-5
Sessional Test – 2	90	20		29.02.2016 to 05.03.2016	Block-5
Sessional Test – 3	90	20		11.04.2016 to 16.04.2016	Block-5
Comprehensive Quiz Examination	20	10	10%	11.04.2016 to 16.04.2016	Block-5
Comprehensive Exam	180	70	70%	25.04.2016 to 07.05.2016	Block-5

Chamber Consultation Hour: 4.00 PM

Venue : CSE Staff Room

Notices: CSE Main Notice Board

Signature of the Instructors

Signature of the Course-Coordinator