



SWARNANDHRA

College of Engineering & Technology (Autonomous)

Narsapur - 534 280.

DEPARTMENT OF COMPUTER & SCIENCE ENGINEERING

DATAMINIG AND DATA WAREHOUSING LESSON PLAN

Course Code	Course Title	Semester	Branch	Contact Hrs/Week	Sections
16CS7T03	Data Mining And Data WareHousing	VII	CSE	6	A,B&C

COURSE OUTCOMES:

CO1: Understand why there is a need for data warehouse in addition to traditional operational Database systems

CO2: Identify components in typical data warehouse architectures

CO3: Process raw data to make it suitable for various data mining Algorithms.

CO4: Design a data warehouse and understand the process required to construct one

CO5: Solve real data mining problems by using the right tools to find interesting patterns

Week No	Outcome	Topics/Activity	Ref . Te xt Bo ok	Con. Hr	Delivery method	
UNIT – I :INTRODUCTION TOTAL HOURS: 10						
1	CO1: Understand why there is a need for data warehouse in addition to traditional operational database systems	1.1	What Motivated Data Mining, Why Is It Important	T2	1	
		1.2	Data Mining—On What Kind of Data	T2	1	
		1.2.1	Relational databases	T2	1	
		1.2.2	Data Warehouses	T2		
		1.2.3	Transactional Databases	T2		
2			1.2.4	Advanced Data and Information Systems and Advanced Applications	T2	
			1.3	Data Mining Functionalities	T2	1
			1.3.1	Characterization and discrimination	T2	
			1.3.2	Mining frequent Patterns,Associations,Correlations	T2	

3	CO5: Solve real data mining problems by using the right tools to find interesting pattern	1.3.3	Classification and prediction	T2	1	
		1.3.4	Cluster analysis	T2		
		1.3.5	Outlier Analysis	T2		
		1.3.6	Evolution Analysis	T2		
		1.4	Are All of the Patterns Interesting?	T2	1	
		1.5	Classification of Data Mining Systems	T2	1	
		1.6	Data Mining Task Primitives	T2	1	
		1.7	Integration of a Data Mining System with a Database or Data Warehouse System	T2	1	
		1.8	Major Issues in Data Mining.	T2	1	

ADD ON TOPICS: Grid DB

Assignment questions:

- 1) Discuss knowledge discovery process in data mining
- 2) List out major issues in data mining
- 3) Discuss about Data mining functionalities.

Model Paper Questions:

- 1) Write about classification of data mining systems(K2)
- 2) List out the five primitives for specifying a data mining task(k1)

UNIT – II: DATA PRE-PROCESSING TOTAL HOURS:10

4 5 6	CO5. Solve real data mining problems by using the right tools to find interesting patterns	2.1	Why Pre-process the Data	T2	1	Chalk and talk PPT Smart Board
		2.2	Descriptive Data summarization	T2	1	
		2.2.1	Measuring the Central Tendency	T2	1	
		2.2.2	Measuring the Dispersion of Data	T2		
		2.2.3	Graphic Display of Basic Description of Data Summaries	T2	1	
		2.3	Data Cleaning	T2	1	
		2.3.1	Missing Values	T2		
		2.3.2	Noisy Data	T2		
		2.3.3	Data cleaning as a Process	T2		
		2.4	Data Integration and Transformation	T1	1	

		2.5	Data Reduction	T2	1	
		2.5.1	Data Cube Aggregation	T2		
		2.5.2	Attribute Subset Selection	T2	1	
		2.5.3	Dimensionality Reduction	T2	1	
		2.5.4	Numerosity reduction	T2		
		2.5.2	Data Discretization and Concept Hierarchy Generation	T2	1	

ADD ON TOPICS: DATA PREPROCESSING IN WEB USAGE MINING

Assignment questions:

- 1) Explain Various data cleaning techniques in pre-processing(k4)
- 2) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order)
13, 15, 16, 16, 19,20,20,21,22,22,25,25,25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52,70.
Compute the following: (K3)
 - a)find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
 - b)Give the five-number summary of the data.
 - c)Show a boxplot of the data.. Discuss various coupling techniques to integrate data mining systems with data warehouse.(k2)

Model Paper Questions:

- 1) Explain Concept hierarchy generation for numerical data (k3)
- 2) Calculate Chi-Square value from the following data samples (K4)

	male	female
Fiction	250	200
Non fiction	50	1000

UNIT – III DATA WAREHOUSE AND OLAP TECHNOLOGY TOTAL HOURS:09

7	CO2: Identify components in typical data warehouse architectures	3.1	What Is a Data Warehouse?	T2	1	Chalk and talk/ PPT Smart Board
		3.2	A Multidimensional Data Model	T2	1	
		3.3	Data Warehouse Architecture	T2	1	
8	CO4: Design a data warehouse and understand the process required to construct one	3.3.1	Steps for the Design and Construction of Data Warehouses	T2	1	
		3.3.2	A Three-Tier Data Warehouse Architecture	T2		
		3.3.3	Data Warehouse Back-End Tools and Utilities	T2	1	

		3.3.4	Metadata Repository	T2	1	
		3.3.5	Types of OLAP servers	T2	1	
		3.4	Data Warehouse Implementation	T2	1	
		3.5	From Data Warehousing to Data Mining	T2	1	

ADD ON TOPICS: DATA LAKE

Assignment questions:

- 1) Define Data warehouse. Differentiate between OLTP and OLAP(K1)
- 2) Explain the Three Tier Data warehouse Architecture with neat diagram.(K3)

Model Paper questions:

- 1) Discuss different OLAP operations with an example(K2)
- 2) Describe about materialization Methods in data warehouse implementation(K1)

9

MID –I EXAMINATIONS

UNIT – IV : CLASSIFICATION

TOTAL HOURS:12

10	CO3. Process raw data to make it suitable for various data mining Algorithms.	4.1	Basic Concepts	T1	1	Chalk and talk/ PPT Smart Board
		4.2	General Approach to solving a classification problem	T1		
		4.3	Decision Tree Induction	T1	1	
		4.3.1	Working of Decision Tree	T1	1	
11	CO5: Solve real data mining problems by using the right tools to find interesting patterns	4.3.2	Building a decision tree	T1	1	
		4.3.3	Methods for expressing an attribute test conditions	T1	1	
		4.3.4	Measures for selecting the best split	T1	1	
		4.3.5	Algorithm for decision tree induction	T1	1	
12		4.4	Model Over fitting	T1	1	
		4.4.1	Due to presence of noise	T1		
		4.4.2	Due to lack of representation samples,	T1	1	
		4.5	Evaluating the performance of classifier:	T1	1	
		4.5.1	Holdout method	T1	1	
		4.5.2	Random subsampling	T1		
		4.5.3	Cross-validation	T1	1	
		4.5.4	Bootstrap.	T1		

ADD ON TOPICS: SPRINT: A Scalable Parallel Classifier for Data Mining

Assignment questions:

1) Explain how to build a decision tree(K2)

2) write about general approach to solving a classification problem(K3)

Model Paper questions:

Describe the different measure for selecting the best split in decision tree induction(K1)

Define Hunts algorithm. How is it helpful to construct decision tree.(k1)

UNIT –V : ASSOCIATION ANALYSIS: BASIC CONCEPTS AND ALGORITHMS -TOTAL HOURS:09

13	<p>CO3. Process raw data to make it suitable for various data mining Algorithms.</p> <p>CO5. Solve real data mining problems by using the right tools to find interesting patterns</p>				
		5.1	Introduction,	T1	1
5.2		Frequent Itemset Generation	T1	1	
5.2.2		Apriori Principle	T1	1	
5.2.3		Frequent Itemset Generation in the Apriori Algorithm	T1	1	
5.2.4		Candidate Generation and pruning	T1	1	
5.2.5		Support Counting	T1	1	
5.2.6		Computational Complexity	T1	1	
5.2.7		Rule generation	T1	1	
5.3		Compact representation of frequent item sets	T1	1	
5.4	FP-Growth Algorithm	T1	1		

ADD ON TOPICS: A Novel Quantity based Weighted Association Rule Mining.**Assignment questions:**

1) Explain FP-growth algorithm for the generation of frequent itemsets(K2)

2) Write the Apriori Algorithm for Rule Generation(K3)

Model Paper questions:

1) Compare Apriori and FP-Growth Algorithm(K4)

2) Write the steps to construct an FP-Tree with an Example(K3)

UNIT –VI : CLUSTER ANALYSIS: BASIC CONCEPTS AND ALGORITHMS TOTAL HOURS:10

15	<p>CO3. Process rw data to make it suitable for various data mining Algorithms.</p> <p>CO5 : Solve real data mining problems by using the right tools to find interesting patterns</p>	6.1	Overview	T1	1	Chalk and talk/ PPT Smart Board
		6.1.1	What Is Cluster Analysis	T1		
		6.1.2	Different Types of Clusterings	T1		
6.1.3		Different Types of Clusters	T1	1		
6.2		K-Means	T1	1		
16		6.2.1	The Basic K-means Algorithm	T1	1	
		6.2.2	K-means Additional Issues	T1		
17		6.2.3	Bisecting K-means	T1	1	
		6.2.4	K-means and Different Types of Clusters	T1		
		6.2.5	Strengths and Weaknesses	T1		

		6.2.6	K-means as an Optimization Problem	T1		
		6.3	Agglomerative Hierarchical Clustering	T1	1	
		6.3.1	Basic Agglomerative Hierarchical Clustering Algorithm	T1	1	
		6.3.2	Specific Techniques	T1		
		6.4	DBSCAN	T1	1	
		6.4.1	Traditional Density	T1		
		6.4.2	The DBSCAN Algorithm	T1	1	
		6.4.3	Strengths and Weaknesses	T1		

ADD ON TOPICS: Text Mining, Graph Mining

	<p>Assignment questions: 1) Describe the strengths and Weaknesses of DBSCAN clustering algorithm(K3) 2) Write basic K-means Algorithm and also discuss additional issues in K-means.(K2) Model Paper Questions: 1) Discuss Agglomerative Hierarchical Clustering Algorithm(K2) 2) Explain different types of clustering methods(K3)</p>					
18	MID –II EXAMINATION					
19	END SEMESTER EXAMINATIONS					
20-21	END SEMESTER EXAMINATIONS					
LECTURE HOUR TOTAL: 60						
ADD ON TOPICS: 06						
GRAND TOTAL: 66						

TEXT BOOKS

S. No	AUTHORS, BOOK TITLE ,EDITION, Publisher, YEAR OF PUBLICATION
1.	Pang-Ning Tan & Michael Steinbach, Vipin Kumar, Introduction to Data Mining Pearson-2016
2.	Jiawei Han, Michel Kamber, Data Mining concepts and Techniques, Elsevier, Third Edition-2011
3.	Arun K Pujari ,Data Mining Techniques, University Press E-edition:First Published 2013
Web Details	
1.	http://www.kdd.org/
2.	nptel.ac.in/courses/106106093/35

	NAME	SIGNATURE WITH DATE
I	FACULTY	N.TULASI RAJU
II	FACULTY	Dr.P.SRINIVASULU
III	COURSE COORDINATOR	N.TULASI RAJU
IV	MODULE COORDINATOR	Dr.P.SRINIVASULU
V	PROGRAMME COORDINATOR	Dr.P.SRINIVASULU

Faculty

Head of the Department

Principal