SREENIVASA INSTITUTE of TECHNOLOGY and MANAGEMENT STUDIES
(AUTONOMOUS)

(DATA MINING AND WAREHOUSING)

QUESTION BANK

III - B.TECH / II - SEMESTER REGULATION: R16

COMPILED BY
FACULTY INCHARGE : MRS.S.KOKILA&MR.KSATHISH
DESIGNATION : ASST PROFESSOR
DEPARTMENT :CSE

.

III B. TECH II-SEMESTER (CSE)

L	T	P	C
3	1	0	3

16CSE321 DATA MINING AND WAREHOUSING

Course Educational Objectives:

The student should be made to:

CEO1: Gain the knowledge aboutthe basics of data mining and data warehousing concepts, data mining functionalities, and Preprocessing concepts.
CEO1: Understand and implement the data warehouse architecture, different data warehouse schemas
CEO1: Analyze and implement the Association Rules for analyzing the Transactional Databases
CEO1: Study and Implement the major Classification and Clustering Algorithms
CEO1: Study the advanced data mining concepts.

UNIT -1:
Introduction: Motivation and Importance of Data Mining - Data Mining - Kind of Data to be mined - Data Mining Functionalities - Kind of patterns to be mined - Classification of Data Mining Systems - Integration of a Data Mining System with a Database or Data Warehouse System - Major Issues in Data Mining.
Data Pre-processing: The need for Preprocessing - Data Cleaning - Data Integration and Transformation - Data Reduction - Data Discretization and Concept Hierarchy Generation.

UNIT -2:
An Overview: Data warehouses and its Characteristics - Data warehouse Architecture and its Components - Extraction - Transformation – Loading - Schema Design - Star and Snow - Flake Schema - Fact Constellation Schema - OLAP Cube - OLAP Operations - OLAP Server Architecture - Data Warehouse Implementation - From Data Warehousing to Data Mining.

UNIT -3:
Association Rules: problems Definition - Frequent Item Set Generation - The APRIORI Principle - Support and Confidence Measures - Association Rule Generation - APRIOIRI Algorithm - The Partition Algorithms - FP - Growth Algorithms - Compact Representation of Frequent Item set - Maximal Frequent Item Set - Closed Frequent Item Sets.

UNIT -4:
Classification: Problem Definition - Evaluation of classifiers - Classification Techniques, Decision Tree - Decision tree Construction - Methods for Expressing attribute test conditions - Measures for Selecting the Best Split - Algorithm for Decision tree Induction - Naive Bayes Classifier - Bayesian Belief Networks – K - Nearest neighbor classification.

Clustering: Clustering Overview - Partitioning Clustering - K-Means and K-Medoids Algorithms - Hierarchical Clustering - Agglomerative Methods and divisive methods - Outlier Detection.

UNIT -5:

Advanced Mining: Multimedia Data Mining - Text Mining - Mining the World Wide Web -Data Mining Applications - Social Impacts of DataMining.

Course Outcomes:

On Successful completion of this course student will be able to :

Course Outcomes		POs related to COs
CO1	Understand the need and applications of Data Warehouse and Data Mining	PO1, PO2
CO2	Designand Implement the data warehouse by using major data warehouse schemas	PO1, PO3
CO3	ImplementAssociation Rules for analyzing Transactional databases	PO1, PO4
CO4	Understand and Implement major Classification And Clustering Algorithms	PO1, PO4
CO5	Apply the Data mining techniques in real time problems.	PO1, PO2, PO4

Text books:

1. Data Mining – Concepts and Techniques, 2 Edition, Jiawei Han, MichelineKamber, 2006, Morgan Kaufmann Publishers, Elsevier.

Reference Books:

1. Data Mining Techniques, 3rd Edition ,Arun K Pujari,Universities Press.
2. Data Warehouse Fundamentals, PualrajPonnaiah, Wiley Student Edition.
3. Data Mining, VikaramPudi, P Radha Krishna, Oxford University Press

QUESTION BANK

Question No.	Questions	PO Attainment
UNIT – 1: Data Mining and Data Preprocessing		
PART-A (Two Marks Questions)		
1	Define Data Mining	PO1
2	List out KDD process steps	PO1
3	What are the types of data?	PO1
4	Compare descriptive and predictive data mining	PO1
5	What is classification	PO1
6	What is prediction	PO1, PO2
7	Why we need to Pre-process the data	PO1
8	List out Data Pre-processing steps	PO1
9	What is Data cleaning	PO1
10	what is Data integration	PO1
11	Illustrate Data transformation functions	PO1
12	List out the major issues in data mining	PO1
13	What is Data selection	PO1
14	Define Data warehouse	PO1
15	Define Outlier Analysis	PO1, PO2
16	Define Clustering analysis	PO1
17	Define evolution Analysis	PO1 ,PO2
18	What is data redundancy	PO1
19	Define Data discretization	PO1
20	What is categorical attribute	PO1
PART-B (Ten Marks Questions)		
1	Identify the need for Data Mining	PO1, PO2
2	Show with diagrammatic illustration of the steps involved in the process of the Knowledge Discovery from Data (10M)	PO1, PO2, PO4
3	(a)Classify the different types of data on which Mining can be performed (5M) (b)Illustrate the architecture of a typical Data mining system (5M)	PO1, PO2
4	Explain Various Data Mining Functionalities with an example (10M)	PO1, PO2
5	(a)Illustrate with a diagram about Data Mining Task Primitives. (b)Discuss about the Major issues in Data Mining.	PO1, PO2
6	What is Data Cleaning? Describe various methods of Data Cleaning.	PO1, PO2
7	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, 23, 29, 35, 41, 44, 53, 62, 69, 72 (i) Use min-max normalization to transform the value of 45 for age onto the range [0,1] (ii) Use Z-Score normalization to transform the value 45 for age where the standard deviation of age is 20.64 years	PO1, PO2, PO4
8	(a)List the Issues to be considered during Data Integration (4M) (b)Discuss about detecting data redundancy using correlation analysis (6M)	PO1, PO2, PO4

9	Explain about Data Transformation method with suitable example	**PO1, PO2, PO4**
10	Explain about the different Data Reduction techniques.	**PO1, PO2, PO4**

Question No.	Questions	PO Attainment
UNIT – 2: Data Warehouse and OLAP Technology		
PART-A (Two Marks Questions)		
1	List the key words used in the definition of Data Warehouse.	**PO1**
2	Define Star Schema.	**PO1**
3	Define Snow flake Schema	**PO1**
4	Define fact Constellation Schema	**PO1**
5	What is concept hierarchy	**PO1**
6	Point out the major differences between the star schema and the snowflake schema.	**PO1**
7	Differentiate between MOLAP and ROLAP.	**PO1**
8	A Data Cube C, has n dimensions and each dimension has exactly P distinct values what is the maximum, minimum Number of cells possible in the base cuboid	**PO1**
9	List the types of OLAP.	**PO1**
10	List the types of Indexing.	**PO1**
11	What is data discretization?	**PO1**
12	Compare the size of Database in OLTP and OLAP	**PO1**
13	Define metadata	**PO1**
14	List out the OLAP Operations	**PO1, PO2**
15	Define star net Query model	**PO1**
16	Define partial materialization	**PO1, PO2**
17	Define full materialization	**PO1**
18	List out the components of data warehousing	**PO1**
19	Define data cube	**PO1**
20	Differentiate between base cuboid and apex cuboid	**PO1**
PART-B (Ten Marks Questions)		
1	Explain about the multidimensional data model.	**PO1, PO2, PO4**
2	Differentiate OLTP and OLAP with features.	**PO1, PO2, PO4**
3	Discuss about OLAP operations in the multidimensional data model	**PO1, PO2, PO4**
4	Explain about schemas in multi dimensional databases	**PO1, PO2, PO4**
5	Explain the three- tier Data Warehouse Architecture with neat diagram	**PO1, PO2, PO4**
6	(a)Discuss about concept hierarchies with suitable example. (b)Discuss about Star-net Query model for querying multidimensional data model.	**PO1, PO2, PO4**
7	Discuss about meta data repository and Back-end tools used in Data Warehouse Architecture.	**PO1, PO2, PO4**

8	(a)Compare the types of OLAP Servers (b)Discuss about efficient computation of Data Cubes.	**PO1, PO2, PO4**
9	Explain about Indexing OLAP Data with an example.	**PO1, PO2, PO4**
10	Explain about from On-Line Analytical Processing to Online Analytical Mining	**PO1, PO2, PO4**

Question No.	Questions	PO Attainment
	UNIT – 3: Association Rules	
	PART-A (Two Marks Questions)	
1	What is meant by association rule?	PO1
2	What is meant by Market basket analysis?	PO1
3	state and explain Apriori property.	PO1, PO2
4	What is meant by Mining Multilevel Association Rules?	PO1
5	Define Uniform Minimum Support.	PO1
6	What is meant by Reduced Minimum Support?	PO1
7	What is meant by multidimensional association rules?	PO1
8	What is meant by intradimensional association rule?	PO1
9	What is meant by inter dimensional association rules?	PO1
10	What is meant by Quantitative association rules?	PO1
11	What is meant by Partition Algorithms?	PO1, PO2
12	state and explain FP_growth Algorithm.	PO1, PO2
13	What is meant by Frequent itemset.	PO1
14	What is meant by Maximal Frequent Item Set?	PO1
15	What is meant by Closed Frequent Item Set?	PO1
16	Expalin the join & prune step in apriori algorithm.	PO1
17	Draw and explain the conditional FP_Tree.	PO1
18	How will you measure support and confidence with an example?	PO1
19	How to improve the efficiency of apriori algorithm.	PO1
20	What is meant by conditional pattern base?	PO, PO2
	PART-B (Ten Marks Questions)	
1	Explain about Various kinds of Association rule Mining.	PO1, PO2
2	Discuss about FP-growth algorithm for the following given example{M,O,N,K,E,Y} {D,O,N,K,E,Y} {M,A,K,E} {M,U,C,K,Y} {C,O,O,K,I,E}, Support= 60 %, Confidence = 80 %.	PO1, PO2, PO4
3	State and explain Apriori Algorithm with an example Consider the following data set to generate Association rules {M,O,N,K,E,Y} {D,O,N,K,E,Y} {M,A,K,E} {M,U,C,K,Y} {C,O,O,K,I,E}, Support= 60 %, Confidence = 80 %.	PO1, PO2, PO4
4	Explain in detail about partitional algorithms with an example.	PO1, PO2
5	Explain the steps involved in Apriori Algorithm.	PO1, PO2, PO4
6	Write short notes on Maximal Frequent Item Set &Closed Frequent Item Set.	PO1, PO2
7	Explain in detail about support and Confidence Measures with an example.	PO1, PO2, PO4
8	Discuss about Quantitative association mining.	PO1, PO2
9	Explain in detail about Multidimensional association rule.	PO1, PO2, PO4
10	State and explain apriori algorithm .for the following given example. Support= 60 %, Confidence = 80 %.	PO1, PO2, PO4

transaction ID	items
1	{A,C,D}
2	{B,C,E}
3	{A,B,C,E}
4	{B,E}
5	{A,B,C,E}

Question No.	Questions	PO Attainment
	UNIT – 4: Classification and Clustering	
	PART-A (Two Marks Questions)	
1	Where are decision trees mainly used?	PO1, PO2
2	What do you meant by concept hierarchies?	PO1
3	How will you solve a classification problem using decision trees?	PO1
4	Explain ID3.	PO1
5	What is a "decision tree"?	PO1, PO2
6	Define Data Classification.	PO1
7	Define Prediction.	PO1
8	What is the difference between "supervised" and unsupervised" learning scheme.	PO1
9	What is clustering?	PO1
10	What are the requirements of clustering?	PO1
11	State the categories of clustering methods?	PO1
12	What do you meant by Bayesian Classification.	PO1
13	State and explain Bayes Theorem..	PO1, PO2
14	Difference between K-Means and K-Medoids Algorithms.	PO1
15	What do you meant by Hierarchical Clustering	PO1
16	What do you meant by Agglomerative Clustering.	PO1
17	What do you meant by Outlier Detection.	PO1
18	What do you meant by divisive Clustering.	PO1
19	What is Bayesian Belief Networks.	PO1
20	What do you meant by best split.	PO1, PO2
	PART-B (Ten Marks Questions)	
1	Discuss about Decision tree induction algorithm with an example.	PO1, PO2, PO4
2	Explain about Attribute Subset Selection Measures with an example.	PO1, PO2, PO4
3	Expalin the Naive Bayesian Classification algorithm.	PO1, PO2, PO4
4	Write short notes on Bayesian Belief Networks?	PO1, PO2, PO4
5	Discuss about k-nearest neighbor classification algorithm with an example?	PO1, PO2, PO4
6	What do you meant by Clustering? Explain the requirements used in Clustering?	PO1, PO2
7	Expain in detail about Hierarchical Clustering.	PO1, PO2
8	Explain in detail about partitional Clustering method.	PO1, PO2
9	Discuss about Outlier Detection.	PO1, PO2
10	Explain in detail about Clustering methods with an example.	PO1, PO2

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES

(Autonomous)

DEPARTMENT of COMPUTER SCIENCE & ENGINEERING

QUESTION BANK **DATA MINING AND WAREHOUSING (16CSE321)**

Question No.	Questions	PO Attainment
UNIT – 5: Advanced Mining and its applications		
PART-A (Two Marks Questions)		
1	What are the foundations of data mining	PO1
2	Name some specific application oriented databases	PO1
3	What are the goals of time series analysis	PO1
4	Name some conventional visualization techniques	PO1, PO2
5	How can data visualization help in decision making	PO1
6	What is the use of DB Miner	PO1, PO2
7	What is the scope of data mining	PO1
8	Differentiate between data mining and data warehousing	PO1
9	Which problems in general the data mining can solve	PO1
10	Short notes on data mining for Financial data analysis	PO1
11	List out the types of data mining	PO1
12	What is meant by information retrieval	PO1
13	What is meant by Content based retrieval system	PO1
14	What is meant by Similarity search in multimedia	PO1
15	List out the techniques in audio mining	PO1
16	List out the techniques in video mining	PO1
17	What is meant by graph based mining	PO1
18	What is meant by network based mining	PO1
19	Mention the types of information retrieval systems	PO1
20	What is meant by text mining	PO1
PART-B (Ten Marks Questions)		
1	Explain how data mining is used in banking industry	PO1, PO2
2	Explain the data mining applications for retail industry	PO1, PO2
3	Explain how data mining is used in health care analysis	PO1, PO2
4	Explain data mining applications for bio medical and DNA data Analysis	PO1, PO2
5	What are the social impacts of Data Mining	PO1, PO2
6	How to mine the world wide web	PO1, PO2
7	Explain briefly how data is analyzed by data mining in Finance Sector	PO1, PO2
8	Discuss about Multimedia data mining	PO1, PO2
9	Discuss about Spatial Data Mining	PO1, PO2
10	Discuss about Text Mining	PO1, PO2

*****ALL THE BEST*****