SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY :: PUTTUR
Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code : Data Warehousing and Data Mining(16CS531)**Course & Branch**: B.Tech.-CSE

Year &Sem: IV B.Tech. &ISem **Regulation:** R16

UNIT –I

1.	Define Data mining? Explain about data mining on what kind of data?	[L2] [CO1]	[12M]
2.	a) What is KDD? Explain about data mining as a step in the process of knowledge discovery	[L2] [CO1]	[6M]
	b) How to classify data mining systems? Explain	[L2] [CO1]	[6M]
3.	Explain about the following a) What motivated Data mining? Explain	[L1] [CO1]	[6M]
	b) Data mining as a step in the process of knowledge discovery	[L2] [CO1]	[6M]
4.	Discuss about Data Mining Task primitives with examples?	[L6] [CO1]	[12M]
5.	Explain in detail about Data mining functionalities?	[L2] [CO1]	[12M]
6.	Describe about Major issues in Data mining?	[L3] [CO1]	[12M]
7.	a) Why do we preprocess the data? Explain.	[L2] [CO1]	[6M]
	b) Write in brief about Data cleaning?	[L2] [CO1]	[6M]
8.	Explain the following?a) Data Integration	[L5] [CO1]	[6M]
	b) Data Transformation methods	[L5] [CO1]	[6M]
9.	What is Data reduction? Discuss in detail?	[L6] [CO1]	[12M]
10.	a) Describe about Data discretization?	[L2] [CO1]	[6M]
	b) Write about Dimensionality reduction methods?	[L2] [CO1]	[6M]

UNIT –II

1.	a) Define Data warehouse? Explain Design principles.	[L2] [CO1]	[6M]
	b) Write in brief about schemas in multidimensional data model	[L2] [CO1]	[6M]
2.	Explain about the Three-tier data warehouse architecture with a neat diagram	[L5] [CO1]	[12M]
3.	Explain about the different types of data models available in DWDM.	[L5] [CO1]	[12M]
4.	a) What are steps in designing the data warehouse? Explain	[L5] [CO1]	[6M]
	b) Compare OLTP and OLAP	[L4] [CO1]	[6M]
5.	Discuss in brief about Data warehouse implementation	[L6] [CO1]	[12M]
6.	Draw and Explain about OLAM Architecture?	[L5] [CO1]	[12M]
7.	Explain in detail about Attribute Oriented Induction with example	[L5] [CO1]	[12M]
8.	a) Briefly explain about the types of OLAP servers.	[L2] [CO1]	[6M]
	b) Explain about Indexing of OLAP data	[L2] [CO1]	[6M]
9.	Discuss in detail about the case study of Data Warehouse for the Government of Andhra Pradesh	[L5] [CO3]	[12M]
10.	Explain the following in OLAP a) Roll up operation	[L2] [CO1]	[3M]
	b) Drill Down operation	[L2] [CO1]	[3M]
	c) Slice operation	[L2] [CO1]	[2M]
	d) Dice operation	[L2] [CO1]	[2M]
	e) Pivot operation	[L2] [CO1]	[2M]

UNIT–III

1. Explain about the Apriori algorithm for finding frequent itemsets with an example [L5] [CO2] [12M]

2. You are given the transaction data shown in the Table below from a fast food restaurant. There are 9 distinct transactions (order:1 – order:9). There are a total of 5 meal items that are involved in the transactions. Forsimplicity we assign the meal items short names (M1 – M5) rather than thefull descriptive names (e.g., Big Mac). [L6] [CO2] [12M]

Meal Item	List of Item IDs	Meal Item	List of Item IDs
Order: 1	M1, M2, M5	Order: 6	M2, M3
Order: 2	M2, M4	Order: 7	M1, M3
Order: 3	M2, M3	Order: 8	M1, M2, M3, M5
Order: 4	M1, M2, M4	Order: 9	M1, M2, M3
Order: 5	M1, M3		

Compute the following :

a) Apply the Apriori algorithm to the dataset of transactions and identify all frequent itemsets with minimum support count as 2. Clearly explain the step by step procedure of the algorithm.

b) Find all strong association rules of the form: $X \Lambda Y \rightarrow Z$ and note their confidence values.

3. a) Explain about basic concepts of frequent itemset mining [L2] [CO1] [6M]

   b) Write the Apriori Algorithm. [L2] [CO2] [6M]

4. a) Write short notes of Multidimensional Association rule [L2] [CO1] [6M]

   b) Discuss about the FP Growth Algorithm. [L5] [CO2] [6M]

5. Let's look at a concrete example, based on the *AllElectronics* transaction database,*D*, of below Table. There are nine transactions in this database, that is, *D* = 9. [L6] [CO2] [12M]

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Solve the problem using Apriori algorithm for finding frequent itemsets in *D*.

6. Describe the steps involved in improving the efficiency of the Apriori algorithm. [L2] [CO3] [12M]

7. Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with an example [L5] [CO2] [12M]

8. What are the Draw backs of Apriori Algorithm? Explain about FP Growth Concept in Detail? [L5] [CO2] [12M]

9. Explain about the Mining Multilevel Association rules with example [L5] [CO2] [12M]

10. a) Write about basic concept in Association Rule Mining [L2] [CO2] [6M]

    b) Can we overcome the draw backs of Apriori algorithm? Explain [L2] [CO2] [6M]

11. What are the various Constraints in Constraint based Association rule mining? Explain [L5] [CO2] [12M]

UNIT–IV

1.	Describe the data classification process with a neat diagram. How does the Naive Bayesian classification works? Explain	[L3] [CO1]	[12M]
2.	Explain decision tree induction algorithm for classifying data tuples and discuss suitable example.	[L5] [CO2]	[12M]
3.	How does the Naïve Bayesian classification works? Explain in detail.	[L6] [CO2]	[12M]
4.	a) What is Bayesian belief network? Explain in detail.	[L2] [CO2]	[6M]
	b) Write a note attribute selection measures.	[L2] [CO1]	[6M]
5.	Explain in detail about Attribute Selection methods in Classification	[L5] [CO1]	[12M]
6.	a) What is Bayes theorem? Explain	[L2] [CO2]	[6M]
	b) Discuss about Naïve Bayesian Classification.	[L6] [CO2]	[6M]
7.	a) What is Bayesian belief network? Explain in detail.	[L2] [CO2]	[6M]
	b) Write a note attribute selection measures	[L1] [CO1]	[6M]
8.	Describe in detail about Rule based Classification.	[L4] [CO2]	[12M]
9.	Write and explain about Classification by Back propagation Algorithm.	[L6] [CO2]	[12M]
10.	a) What is prediction? Explain about Linear regression method	[L5] [CO2]	[6M]
	b) Discuss about Accuracy and Error measures.	[L4] [CO1]	[6M]

UNIT –V

1.	Define Clustering? Explain about Types of Data in Cluster Analysis?	[L2] [CO1]	[12M]
2.	a) Classify various Clustering methods.	[L3] [CO1]	[6M]
	b) Explain any one Partitioning based clustering methods.	[L5] [CO2]	[6M]
3.	What is the goal of clustering? How does partitioning around medoids algorithm achieve this goal?	[L4] [CO2]	[12M]
4.	a) Differentiate between AGNES and DIANA algorithms.	[L3] [CO2]	[6M]
	b) How to access the cluster quality?	[L1] [CO1]	[6M]
5.	a) What is outlier detection? Explain distance based outlier detection.	[L2] [CO2]	[6M]
	b) Write partitioning around mediods algorithm.	[L5] [CO2]	[6M]
6.	a) Write K-means clustering algorithm.	[L2] [CO2]	[6M]
	b) Write the key issue in hierarchical clustering algorithm.	[L2] [CO2]	[6M]
7.	Explain the following a) Density based clustering methods b) Grid based clustering methods	[L2] [CO2] [L2] [CO2]	[6M] [6M]
8.	What are outliers? Discuss the methods adopted for outlier detection.	[L6] [CO1]	[12M]
9.	a) What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem?	[L1] [CO2]	[7M]
	b) Give a brief note on PAM Algorithm	[L2] [CO2]	[5M]
10.	Discuss in detail about Data mining Applications.	[L6] [CO3]	[12M]