



SIDDHARTH GROUP OF INSTITUTIONS ::PUTTUR
Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code: DWDM (16CS531)
Year & SEM: III B.Tech & I Sem

Course & Branch: B.Tech-CSE
Regulation: R16

UNIT –I

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

UNIT -II

1. a) Define Data warehouse? Discuss Design principles. 5M
b) Write in brief about schemas in multidimensional data model. 5M
2. Explain about the Three-tier data warehouse architecture with a neat diagram. 10M
3. Discuss the following
 - a) Star schema 3M
 - b) Snow Flake schema 3M
 - c) Fact constellation schema 4M
4. a) What are steps in designing the data warehouse? Explain 5M
b) Compare OLTP and OLAP 5M
5. Describe in brief about Data warehouse implementation 10M
6. Draw and Explain about OLAM Architecture? 10M
7. Write in detail about Attribute Oriented Induction with example 10M
- 8 a) briefly explain the German super market EDEKA's Data warehouse 5M
b) Write about case study of Data Warehousing in the Tamilnadu Government 5M
9. Discuss in detail about the case study of Data Warehouse for the Government of Andhra Pradesh 10M
10. Explain the following in OLAP
 - a) Roll up operation 2M
 - b) Drill Down operation 2M
 - c) Slice operation 2M
 - d) Dice operation 2M
 - e) Pivot operation 2M

UNIT-III

1. Explain about the Apriori algorithm for finding frequent item sets with an example. 10M
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) and each transaction involves between 2 and 4 meal items. There are a total of 5 meal items that are involved in the transactions. For simplicity we assign the meal items short names (M1 – M5) rather than the full descriptive names (e.g., Big Mac).

10M

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		

For all of the parts below the minimum support is $2/9$ (.222) and the minimum confidence is $7/9$ (.777). Note that you only need to achieve this level, not exceed it. Show your work for full credit (this mainly applies to part a). a. Apply the Apriori algorithm to the dataset of transactions and identify all frequent k itemset. Show all of your work. You must show candidates but can cross them off to show the ones that pass the minimum support threshold. This question is a bit longer than the homework questions due to the number of transactions and items, so proceed carefully and neatly. Note: if a candidate itemset is pruned because it violates the Apriori property, you must indicate that it fails for this reason and not just because it does not achieve the necessary support count (i.e., in these cases there is no need to actually compute the support count). So, explicitly tag the itemset that are pruned due to violation of the Apriori property. This really did not come up on the homework

because those problems were quite short. (If you do not know what the Apriori property is, do not panic. You will ultimately get the exact same answer but will just lose a few points). b. Find all strong association rules of the form: $X \wedge Y \diamond Z$ and note their confidence values. Hint: the answer is not 0 so you should have at least one frequent 3-frequent itemset.

- 3 a) Discuss about basic concepts of frequent itemset mining. 5M
 b) Write the Apriori Algorithm. 5M
- 4 a) what are the drawbacks of Apriori Algorithm? Explain 5M
 b) Write the FP Growth Algorithm. 5M
- 5 a) what are the advantages of FP-Growth algorithm? 5M
 b) Discuss the applications of association analysis. 5M
6. Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with an example 10M
7. What are the Draw backs of Apriori Algorithm? Explain about FP Growth Concept in Detail? 10M
8. Explain about the Mining Multilevel Association rules with example. 10M
- 9 a) Write about basic concept in Association Rule mining 5M
 b) Can we overcome the draw backs of Apriori algorithm? Discuss. 5M
10. What are the various Constraints in Constraint based Association rule mining? Explain. 10M

UNIT-IV

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

UNIT -V

1. Define Clustering? Explain about Types of Data in Cluster Analysis? 10M
2. a) Classify various Clustering methods. 5M
b) Write any one Partitioning based clustering methods. 5M
3. What is the goal of clustering? How does partitioning around medoids algorithm achieve this? 10M
4. a) Differentiate between AGNES and DIANA algorithms. 5M
b) How to access the cluster quality? 5M
5. a) What is outlier detection? Explain distance based outlier detection. 5M

- b) Write partitioning around medoids algorithm. 5M
6. a) Write K-means clustering algorithm. 5M
- b) Write the key issue in hierarchical clustering algorithm. 5M
- 7 Explain the following
- a) Density based clustering methods 5M
- b) Grid based clustering methods 5M
8. What are outliers? Discuss the methods adopted for outlier detection 10M
9. a) Give a brief note on PAM Algorithm. 4M
- b) What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? 6M
10. Discuss in detail about Data mining Applications. 10M