

Code No: 117CD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD**B. Tech IV Year I Semester Examinations, November/December - 2017****DATA WAREHOUSING AND DATA MINING****(Computer Science and Engineering)****Time: 3 Hours****Max. Marks: 75****Note:** This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART - A**(25 Marks)**

- 1.a) Define data warehouse. [2]
- b) List the Data warehouse Characteristics. [3]
- c) How can you go about filling in the missing values for this attribute? [2]
- d) Why is the word data mining a misnomer? [3]
- e) Give a note on Closed Frequent Item Set. [2]
- f) Write the FP-graph algorithm. [3]
- g) How prediction is different from classification? [2]
- h) What is rule classification? [3]
- i) Give a note on k means algorithm. [2]
- j) List the Key Issues in Hierarchical Clustering. [3]

PART – B**(50 Marks)**

- 2.a) Make a comparisons between the MOLAP and HOLAP.
 - b) Discuss the star and snowflake schema in detail with suitable example. [5+5]
- OR**
- 3.a) Write the difference between designing a data warehouse and an OLAP cube.
 - b) Give a brief note on ROLAP. [5+5]
4. Explain concept hierarchy generation for the nominal data. [10]
- OR**
- 5.a) Describe the Feature Subset Selection.
 - b) Illustrate the Data Transformation by Normalization. [5+5]

6. Make a comparison of Apriori and ECLAT algorithms for frequent item set mining in transactional databases. Apply these algorithms to the following data:

TID	LIST OF ITEMS	
1	Bread, Milk, Sugar, TeaPowder, Cheese, Tomato	
2	Onion, Tomato, Chillies, Sugar, Milk	
3	Milk, Cake, Biscuits, Cheese, Onion	
4	Chillies, Potato, Milk, Cake, Sugar, Bread	
5	Bread, Jam, Mik, Butter, Chilles	
6	Butter, Cheese, Paneer, Curd, Milk, Biscuits	
7	Onion, Paneer, Chilies, Garlic, Milk	
8	Bread, Jam, Cake, Biscuits, Tomato	[10]

OR

7. Briefly explain the Partition Algorithms. [10]

8. Discuss K- Nearest neighbor classification-Algorithm and Characteristics. [10]

OR

9. How does the Naïve Bayesian classification works? Explain in detail. [10]

- 10.a) Give a brief note on PAM Algorithm.

- b) What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? [5+5]

OR

11. What are the different clustering methods? Explain in detail. [10]

--ooOoo--

Code No: 117CD

R13

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, March - 2017

DATA WAREHOUSING AND DATA MINING

(Computer Science and Engineering)

Time: 3 Hours

Max. Marks: 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

Part- A (25 Marks)

- 1.a) What is a data mart? [2]
- b) What is a fact table? [3]
- c) What is data mining? [2]
- d) List similarity measures. [3]
- e) What is maximal frequent itemset? [2]
- f) How to compute confidence measure for an association rule? [3]
- g) What is classification? [2]
- h) Define information gain. [3]
- i) What is an outlier? [2]
- j) List the demerits of k-means algorithm. [3]

Part-B (50 Marks)

2. What are the various components of data warehouse? Explain their functionality in detail. [10]

OR

3. What is the significance of OLAP in data warehouse? Describe OLAP operations with necessary diagram/example. [10]

4. Explain different data mining tasks for knowledge discovery. [10]

OR

5. What is the need of dimensionality reduction? Explain any two techniques for dimensionality reduction. [10]

6. A database has six transactions. Let min-sup = 50% and min-conf = 75%.

TID	List of items
001	Pencil, sharpener, eraser, color papers
002	Color papers, charts, glue sticks
003	Pencil, glue stick, eraser, pen
004	Oil pastels, poster colours, correction tape
005	Whitener, pen, pencil, charts, glue stick
006	Colour pencils, crayons, eraser, pen

Find all frequent item sets using Apriori algorithm. List all the strong association rules.

www.ManaResults.co.in

[10]

OR

- 7.a) What are the advantages of FP-Growth algorithm?
b) Discuss the applications of association analysis. [5+5]

8. Explain decision tree induction algorithm for classifying data tuples and discuss suitable example. [10]

OR

- 9.a) What are the characteristics of k-nearest neighbor algorithm?
b) How to evaluate the classifier accuracy? [5+5]

10. What is the goal of clustering? How does partitioning around medoids algorithm achieve this goal? [10]

OR

- 11.a) Differentiate between AGNES and DIANA algorithms.
b) How to access the cluster quality? [5+5]

--ooOoo--

Code No: 117CD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, November/December - 2016

DATA WAREHOUSING AND DATA MINING

(Computer Science and Engineering)

Time: 3 Hours

Max. Marks: 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit.

Each question carries 10 marks and may have a, b, c as sub questions.

PART- A**(25 Marks)**

- 1.a) Define Data ware housing. [2]
- b) Differentiate OLAP, ROLAP and HOLAP. [3]
- c) Discuss about subset selection [2]
- d) Mention any three measures of Similarity. [3]
- e) Define Association rule mining two step processes. [2]
- f) Write short note on support and confidence measures. [3]
- g) Mention types of classifier techniques. [2]
- h) Define Pre pruning and post pruning. [3]
- i) Discuss on Agglomerative and Divisive clustering techniques. [2]
- j) Mention the various types of clustering methods. [3]

PART-B**(50 Marks)**

2. Explain data mining as a step process of knowledge discovery. Mention the Functionalities of Data mining. [10]

OR

3. Differentiate Operational database systems and data warehousing. Explain the star schema and fact constellation schemas. [10]
4. Explain the various Data pre-processing techniques. How data reduction helps in data pre-processing. [10]

OR

5. How can the data cube be efficiently constructed for discovery-driven Exploration? Explain various operations of a Data Cube. [10]
6. How can we mine multilevel Association rules efficiently using concept hierarchies? Explain. Illustrate with an A-priori algorithm for the given dataset below. [10]

TID	List of items
001	milk, dal, sugar, bread
002	Dal, sugar, wheat,jam
003	Milk, bread, curd, paneer
004	Wheat, paneer, dal, sugar
005	Milk, paneer, bread
006	Wheat, dal, paneer, bread

OR

7. Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with example table mentioned above. [10]

8. Describe the data classification process with a neat diagram. How does the Naive Bayesian classification works? Explain. [10]

OR

9. What is prediction? Explain the various prediction techniques. Explain about Decision tree Induction classification technique. [10]

10. What are outliers? Discuss the methods adopted for outlier detection. [10]

OR

11. State K-means algorithm. Apply k-means algorithm with two iterations to form two clusters by taking the initial cluster centers as subjects 1 and 4. [10]

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

--ooOoo--