

**III B. TECH I SEMESTER REGULAR/SUPPLEMENTARY EXAMINATIONS,
NOVEMBER - 2023
DATA WAREHOUSING AND DATA MINING
(Common to CSE & INF BRANCHES)**

Time: 3 hours

Max. Marks: 70

Note : Answer ONE question from each unit (5 × 14 = 70 Marks)

~~~~~

UNIT-I

1. a Explain the three-tier architecture of data warehouse. [7M]  
    )
- b Differentiate between OLTP and OLAP. [7M]  
    )

(OR)

2. a What is data mining? Briefly explain the Knowledge discovery [7M]  
    ) process.
- b What are the major issues in Data Mining? Explain briefly. [7M]  
    )

UNIT-II

3. a Discuss briefly about data cleaning techniques. [7M]  
    )
- b Why do we need data transformation? What are the different ways [7M]  
    ) of data transformation?

(OR)

4. Explain various data pre-processing methods with appropriate [14M]  
    examples. ]

UNIT-III

5. a Explain in detail about decision tree induction algorithm for [7M]  
    ) classifying data tuples with suitable example.
- b What is attribute selection measure? Briefly describe the [7M]  
    ) attribute selection measures for decision tree induction

(OR)

6. Describe the data classification process with a neat diagram. How [14M]  
    does the Naive Bayesian classification works? Explain. ]

UNIT-IV

7. Explain about the Apriori algorithm for finding frequent item sets [14M]  
    with an example. ]

(OR)

8. a What is FP Growth Algorithm? What are the advantages of FP- [7M]

) Growth algorithm?

b Explain how association rules are generated from frequent item sets. [7M]

UNIT-V

9. a Write about K-means clustering algorithm. What is the drawback of k-means algorithm? [7M]

b Write about the key issue in hierarchical clustering algorithm. [7M]

(OR)

10 a Briefly Explain Bisecting K-means algorithm [7M]

b Explain the DBSCAN algorithm. State the pros and cons of DBSCAN method. [7M]

\* \* \* \* \*

**III B. TECH I SEMESTER SUPPLEMENTARY EXAMINATIONS, MAY-2024**  
**DATA WAREHOUSING AND DATA MINING**  
**(COMMON TO CSE AND INF BRANCHES)**

**Time: 3 Hours**  
**70**

**Max. Marks:**

**Note: Answer ONE question from each unit ( $5 \times 14 = 70$  Marks)**

~~~~~

UNIT-I

1. a) What is Data warehouse? What are steps in designing the data warehouse? Explain. [7M]
- b) What is Data Cube? What are the types of Data Cube? Briefly explain the operations performed on Data Cube. [7M]

(OR)

2. a) What are the major issues in Data Mining? Explain briefly. [7M]
- b) Explain the steps involved in the Data Mining Process. Give the sketch of the KDD process. [7M]

UNIT-II

3. a) List different methods for data discretization and explain in detail. [7M]
- b) Explain the pre-processing required to handle missing data and noisy data during the process of data mining. [7M]

(OR)

4. a) List and explain various Data reduction strategies. [7M]
- b) What is noise? Why data smoothing is required. Perform smoothing by bin means, by bin medians and by bin boundaries on the following data. Consider the data for price (in dollars): 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34 [7M]

UNIT-III

5. a) Explain in detail about decision tree induction algorithm for classifying data tuples with suitable example. [7M]
- b) What is association and correlation? With an example describe classification and prediction. [7M]

(OR)

6. a) What are Bayesian classifiers? With an example, describe how to predict a class label using Naive Bayesian classification [7M]
- b) Discuss K- Nearest neighbor classification algorithm and its [7M]

characteristics.

UNIT-IV

7. Make a comparison of Apriori and FP-Growth algorithms for frequent item set mining in transactional databases. Apply these algorithms to the following data: [14M]

TID LIST OF ITEMS

1	Bread, Milk, Sugar, Tea Powder, Cheese, Tomato
2	Onion, Tomato, Chillies, Sugar, Milk
3	Milk, Cake, Biscuits, Cheese, Onion
4	Chillies, Potato, Milk, Cake, Sugar, Bread
5	Bread, Jam, Milk, Butter, Chillies
6	Butter, Cheese, Paneer, Curd, Milk, Biscuits
7	Onion, Paneer, Chillies, Garlic, Milk
8	Bread, Jam, Cake, Biscuits, Tomato

(OR)

8. a) Illustrate the frequent itemset generation using the Apriori algorithm. [7M]
b) Write the algorithm to discover frequent item sets without candidate generation and explain it with an example. [7M]

UNIT-V

9. a) What are the requirements for cluster analysis? Explain briefly. [7M]
b) What is Density based clustering? Describe DBSCAN clustering algorithm. [7M]

(OR)

10. a) What is an outlier? Explain the various types of outliers. [7M]
b) Compare K-Means with Agglomerative Hierarchical clustering. [7M]

* * * * *

III B. TECH I SEMESTER REGULAR EXAMINATIONS, NOVEMBER - 2022
DATA WAREHOUSING AND DATA MINING
(Common to CSE and INF Branches)

Time: 3 Hours

Max. Marks: 70

Note: Answer **ONE** question from each unit (**5 × 14 = 70 Marks**)

UNIT-I

1. a) Suppose that a data warehouse consists of the three [7M]
dimensions time, doctor, and patient, and the two measures
count and charge, where charge is the fee that a doctor
charges a patient for a visit. Draw a schema diagram for the
above data warehouse using star schema.
- b) Describe the steps involved in data mining when viewed as a [7M]
process of knowledge discovery.

(OR)

2. a) Distinguish between OLTP vs OLAP. [7M]
- b) Briefly explain the following data mining functionalities. [7M]
 - i) Association Analysis ii) Classification
 - ii) Cluster Analysis

UNIT-II

3. a) Explain different methods of filling missing values in the [7M]
process of data cleaning.
 - b) Explain the issues to be considered during data integration. [7M]
- (OR)
4. a) What do you mean by attribute subset selection? Discuss [7M]
different techniques used for it.
 - b) With suitable example explain data transformation using [7M]
min-max normalization.

UNIT-III

5. a) With a neat block diagram, explain the general approach to [7M]
solve a classification problem.
 - b) State Bayes theorem. Briefly explain Bayesian classification. [7M]
- (OR)
6. a) What is a decision tree? How it is used to classify an unknown [7M]
data? Explain.
 - b) Write and explain decision tree induction algorithm. [7M]

UNIT-IV

7. a) What are the limitations of Apriori algorithm? Briefly explain. [7M]
 b) With suitable example, explain how to generate association rules from a frequent itemset (Size of itemset should be at least 3). [7M]

(OR)

8. a) Consider the following transactional database. [14M]

TID	Items Bought
T1	{ I ₁ , I ₂ , I ₅ }
T2	{ I ₂ , I ₄ }
T3	{ I ₂ , I ₃ }
T4	{ I ₁ , I ₂ , I ₄ }
T5	{ I ₁ , I ₃ }
T6	{ I ₂ , I ₃ }
T7	{ I ₁ , I ₃ }
T8	{ I ₁ , I ₂ , I ₃ , I ₅ }
T9	{ I ₁ , I ₂ , I ₃ }

Show the FP-tree and the header table. For each frequent item, show how to generate the conditional pattern bases and conditional FP-trees, and the frequent itemsets generated by them.

UNIT-V

9. a) Suppose that the data mining task is to cluster the following eight points (with (x, y) representing location) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9). The distance function is Manhattan distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only. [14M]
 a) The three cluster centers after the first round execution.
 b) The final three clusters.

(OR)

10. a) Illustrate with suitable example k-means clustering algorithm. [7M]
 What are its advantages and disadvantages?
 b) What is density based clustering? Explain DBSCAN algorithm. [7M]

* * * * *

III B. TECH I SEMESTER SUPPLEMENTARY EXAMINATIONS, MAY - 2023
DATA WAREHOUSING AND DATA MINING
(COMMON TO CSE AND INF BRANCHES)

Time: 3 hours

Max. Marks: 70

Note: Answer **ONE** question from each unit (**5 × 14 = 70 Marks**)

~~~~~

UNIT-I

1. a) Illustrate with suitable example the following data warehouse schemas. (i) Star schema (ii) Snowflake schema (iii) Fact constellation schema [7M]  
b) Discuss in details the evolution of database technology. [7M]  
(OR)
2. a) With a neat sketch, briefly explain the three tier architecture of a data warehouse. [7M]  
b) Describe the steps involved in data mining when viewed as a process of knowledge discovery. [7M]

UNIT-II

3. a) 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. Using a bin depth of 3, smooth this data using smoothing by bin means. [7M]  
b) Using the data for age given in question 3 (a), answer the following: [7M]  
i) Use min-max normalization to transform the value 35 for age onto the range [0.0, 1.0].  
ii) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years  
(OR)
4. a) What is sampling? Explain the following sampling techniques used in data reduction. [7M]  
i) Simple random sample without replacement (SRSWOR)  
ii) Stratified sample  
b) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem. [7M]

UNIT-III

5. Build the decision tree using any decision tree algorithm for the concept Play-Golf using the database shown below. Also predict the class label of the unknown data X given below using the decision tree constructed. [14M]

X=(Outlook=Rainy, Temp=cool, Humidity=High, Wind=True

| Outlook  | Temp | Humidity | Windy | Play-Golf |
|----------|------|----------|-------|-----------|
| Rainy    | Hot  | High     | False | No        |
| Rainy    | Hot  | High     | True  | No        |
| Overcast | Hot  | High     | False | Yes       |
| Sunny    | Mild | High     | False | Yes       |
| Sunny    | Cool | Normal   | False | Yes       |
| Sunny    | Cool | Normal   | True  | No        |
| Overcast | Cool | Normal   | True  | Yes       |
| Rainy    | Mild | High     | False | No        |
| Rainy    | Cool | Normal   | False | Yes       |
| Sunny    | Mild | Normal   | False | Yes       |
| Rainy    | Mild | Normal   | True  | Yes       |
| Overcast | Mild | High     | True  | Yes       |
| Overcast | Hot  | Normal   | False | Yes       |
| Sunny    | Mild | High     | True  | No        |

(OR)

6. a) Why do we need to prune the decision tree? Briefly explain the approaches used for tree pruning. [7M]
- b) Why is naïve Bayesian classification called “naïve”? Briefly outline the major ideas of naïve Bayesian classification. [7M]

UNIT-IV

7. a) Distinguish between frequent itemset and closed itemset. [7M]
- b) What is the difference between an association rule and a strong association rule? Explain your answer with a suitable example. [7M]

(OR)

8. Explain with suitable example how Apriori algorithm finds frequent itemsets from the transactional database. [14M]

UNIT-V

9. What is the difference between k-means and k-medoids algorithms? Explain your answer with an example. [14M]

(OR)

10. a) Distinguish between agglomerative and divisive hierarchical clustering. [7M]
- b) State the strengths and weaknesses of Density Based clustering algorithm. [7M]

\* \* \* \* \*