

**III B. TECH II SEMESTER REGULAR EXAMINATIONS APRIL - 2023
DATA WAREHOUSING AND DATA MINING****(CSE - INTERNET OF THINGS, CYBER SECURITY INCLUDING BLOCKCHAIN TECHNOLOGY)**

Time: 3 hours

Max. Marks: 70

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

## UNIT-I

1. a) Analyze the Multi-tiered Architecture of Data Warehouse. [7M]  
b) Differentiate between multidimensional data model. [7M]

(OR)

2. a) Discuss the major issues of Data Mining. [7M]  
b) How is data warehouse different from a database? How are they similar? [7M]

## UNIT-II

3. a) Compare data Integration and data Reduction. [7M]  
b) With an example explain various forms of data preprocessing. [7M]

(OR)

4. a) Classify the various methods for data smoothing. [7M]  
b) Suppose that the data for analysis include the attribute age. The age values for the data tuples are

13,15,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,  
40,45,46,52,60.

Use smoothing by bin means to smooth the above data using a bin depth of 3. Illustrate your steps.

## UNIT-III

5. a) Why naive bayes classification is called naive? Briefly outline the major steps involved in Bayes classification. [7M]  
b) Explain Linear Regression with an suitable example. [7M]

(OR)

6. a) Analyse different attribute selection methods. [7M]  
b) Explain about Bayesian Classification Methods. [7M]

## UNIT-IV

7. Find all frequent item sets for the given training set using Apriori and FP growth respectively. Compare the efficiency of the two mining processes. If the minimum support is 2. [14M]

| TID  | ITEMS BROUGHT           |
|------|-------------------------|
| T100 | {M , O, N , K , E , Y } |
| T200 | {D , O, N, K , E, Y }   |
| T300 | {M , A K, E }           |
| T400 | {M ,U , C , K ,Y }      |
| T500 | {C , O , O ,K , I , E}  |

(OR)

8. a) Explain the Market Basket analysis problem. [7M]  
 b) Explain FP-Growth Algorithm. [7M]

## UNIT-V

9. a) Discuss the steps in K-means algorithm and evaluate the following table using K- means. [7M]

| Subject | A   | B   |
|---------|-----|-----|
| 1       | 1.0 | 1.0 |
| 2       | 1.5 | 2.0 |
| 3       | 3.0 | 4.0 |
| 4       | 5.0 | 6.0 |
| 5       | 3.5 | 5.0 |
| 6       | 4.5 | 5.0 |
| 6       | 3.5 | 4.5 |

- b) Compare between Density based methods and Hierarchical Methods. [7M]

(OR)

10. a) Differentiate AGNES and DIANA hierarchical clustering algorithms. [7M]  
 b) What is meant by cluster analysis. Explain different types of clustering. [7M]

\* \* \* \* \*