



GIET UNIVERSITY, GUNUPUR – 765022

B. Tech (Fifth Semester Regular) Examinations, December – 2023
21BCSPC35002 – Data Mining and Data Warehousing
 (CSE)

Time: 3 hrs

Maximum: 70 Marks

Answer all questions
(The figures in the right hand margin indicate marks)

PART – A

(2 x 5 = 10 Marks)

Q.1. Answer **ALL** questions

- | | CO # | Blooms
Level |
|--|------|-----------------|
| a. Why data transformation is essential in the process of knowledge discovery? | CO1 | K1 |
| b. What smoothing techniques are available to remove noise? | CO1 | K1 |
| c. How are association rules mined from large databases? | CO2 | K1 |
| d. Distinguish between Coverage and Accuracy of a Rule. | CO3 | K4 |
| e. What are the application areas of Data Mining? | CO4 | K1 |

PART – B

(15 x 4 = 60 Marks)

Answer **ALL** questions

- | | Marks | CO # | Blooms
Level |
|--|-------|------|-----------------|
| 2. a. Explain the steps of KDD with the help of a diagram. | 7 | CO1 | K1 |
| b. Find the Chi square correlation analysis for the given four entity instances. | 8 | CO1 | K2 |

Qualification/status	Middle school	High school	Bachelor	Master	Ph.D.
Never married	18	36	21	9	6
Married	12	36	45	36	21
Divorced	6	9	9	3	3
Widowed	3	9	9	6	3

(OR)

- | | | | |
|---|---|-----|----|
| c. What is normalization? Explain why normalization is performed? | 7 | CO1 | K1 |
| d. Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results: | 8 | CO1 | K2 |

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- | | | | |
|--|---|-----|----|
| (a) Calculate the mean, median, and standard deviation of age and %fat. | | | |
| (b) Find out the covariance and correlation among these two attributes. | | | |
| 3.a. Define data warehouse. Draw the architecture of data warehouse and explain the three tiers in detail with a case study. | 8 | CO2 | K1 |

b.	Differentiate between star schema, snowflake schema and fact constellation.	7	CO2	K4
	(OR)			
c.	Distinguish between OLAP and OLTP. List out the various OLAP operations carried out in Data Warehouse.	8	CO2	K4
d.	What is the difference between Virtual data warehouse and enterprise data warehouse?	7	CO2	K1
4.a.	There are five transactions (T1, T2, T3, T4, T5) with items (A, B, C, D) purchased as T1(B, C), T2(A, C, D), T3(B, C), T4(A, B, C, D), T5(B, D). The min_sup = 2. Show how FP-growth approach can generate the association rules for the above dataset.	8	CO3	K2
b.	Explain Decision tree induction algorithm for classification. Discuss the usage of information gain in this.	7	CO3	K2
	(OR)			
c.	Discuss about the attribute selection measures in constructing a decision tree with an example.	8	CO3	K2
d.	Describe KNN Algorithm for data classification with appropriate example.	7	CO3	K2
5.a.	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)	8	CO4	K2
b.	Elaborate the various partitioning methods in detail.	7	CO4	K2
	(OR)			
c.	Differentiate Agglomerative and Divisive Hierarchical Clustering?	8	CO4	K2
d.	How can we use data mining in the field of retail and telecommunication industry?	7	CO4	K2

--- End of Paper ---