



R20 Regulation

Subject code:3P6EB

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

(Autonomous, Accredited by NAAC with 'A+' Grade)

B.Tech VI Semester Supplementary Examinations, December 2024

Data Warehousing and Data Mining (CSE)

Maximum Marks: 70

Date:30.12.2024

Duration: 3 hours

- Note:
- 1.This question paper contains two parts A and B.
 2. Part A is compulsory which carries 20 marks. Answer all questions in Part A.
 3. Part B consists of 5 Units. Answer any one full question from each unit which carries 10M.
 4. Each question carries 10 marks and may have a, b, c, d as sub questions.

Part-A

All the following questions carry equal marks (10X2M=20) Marks)		CO	Bloom Tx
1	List and define the characteristics of data warehouse.	1	L1
2	Define metadata.	1	L1
3	List the methods for filling missing values	2	L1
4	Define the process of KDD	2	L1
5	Mention the importance of association rule mining	3	L1
6	State the property of Apriori	3	L1
7	Write short note on entropy and information gain with formulae	4	L1
8	List the measures for selecting best split in decision tree construction	4	L1
9	List the applications of clustering analysis	5	L1
10	Differentiate clustering and classification	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		CO	Bloom Tx
11	Draw and explain 3-tier architecture of a data warehouse. [10M]	1	L2
OR			
12	List and explain OLAP operations with an example. [10M]	1	L2
13	A) Given the following measurements for the variable age [5M] 18,22,25,42,28,43,33,35,56,28 Standardize the variable by using Z-score. B) Explain data reduction methods. [5M]	2	L2
OR			
14	A) Briefly discuss the data smoothing methods with an example. [5M] B) Explain how to handle missing values with an example. [5M]	2	L2
15	Consider the following transactional data base , generate association rules and identify strong association rules using APRIORI algorithm with support=2 confidence=60% [10M]	3	L2

		<table><tr><td>Tid</td><td>items</td></tr><tr><td>T1</td><td>Hotdogs, bun, ketchup</td></tr><tr><td>T2</td><td>Hotdogs, bun</td></tr><tr><td>T3</td><td>Hotdogs, coke, chips</td></tr><tr><td>T4</td><td>coke, chips</td></tr><tr><td>T5</td><td>Chips,ketchup</td></tr><tr><td>T6</td><td>Hotdogs, coke, chips</td></tr></table>	Tid	items	T1	Hotdogs, bun, ketchup	T2	Hotdogs, bun	T3	Hotdogs, coke, chips	T4	coke, chips	T5	Chips,ketchup	T6	Hotdogs, coke, chips																																																																																					
Tid	items																																																																																																				
T1	Hotdogs, bun, ketchup																																																																																																				
T2	Hotdogs, bun																																																																																																				
T3	Hotdogs, coke, chips																																																																																																				
T4	coke, chips																																																																																																				
T5	Chips,ketchup																																																																																																				
T6	Hotdogs, coke, chips																																																																																																				
	OR																																																																																																				
16	Write and explain FP-growth algorithm with an example.	[10M]	3	L2																																																																																																	
17	Construct a Decision tree using ID3 for below data set	[10M]	4	L2																																																																																																	
	<table><tr><td></td><td>1: outlook</td><td>2: temperature</td><td>3: humidity</td><td>4: windy</td><td>5: play</td></tr><tr><td></td><td>Nominal</td><td>Nominal</td><td>Nominal</td><td>Nominal</td><td>Nominal</td></tr><tr><td>1</td><td>sunny</td><td>hot</td><td>high</td><td>FALSE</td><td>no</td></tr><tr><td>2</td><td>sunny</td><td>hot</td><td>high</td><td>TRUE</td><td>no</td></tr><tr><td>3</td><td>overcast</td><td>hot</td><td>high</td><td>FALSE</td><td>yes</td></tr><tr><td>4</td><td>rainy</td><td>mild</td><td>high</td><td>FALSE</td><td>yes</td></tr><tr><td>5</td><td>rainy</td><td>cool</td><td>normal</td><td>FALSE</td><td>yes</td></tr><tr><td>6</td><td>rainy</td><td>cool</td><td>normal</td><td>TRUE</td><td>no</td></tr><tr><td>7</td><td>overcast</td><td>cool</td><td>normal</td><td>TRUE</td><td>yes</td></tr><tr><td>8</td><td>sunny</td><td>mild</td><td>high</td><td>FALSE</td><td>no</td></tr><tr><td>9</td><td>sunny</td><td>cool</td><td>normal</td><td>FALSE</td><td>yes</td></tr><tr><td>10</td><td>rainy</td><td>mild</td><td>normal</td><td>FALSE</td><td>yes</td></tr><tr><td>11</td><td>sunny</td><td>mild</td><td>normal</td><td>TRUE</td><td>yes</td></tr><tr><td>12</td><td>overcast</td><td>mild</td><td>high</td><td>TRUE</td><td>yes</td></tr><tr><td>13</td><td>overcast</td><td>hot</td><td>normal</td><td>FALSE</td><td>yes</td></tr><tr><td>14</td><td>rainy</td><td>mild</td><td>high</td><td>TRUE</td><td>no</td></tr></table>				1: outlook	2: temperature	3: humidity	4: windy	5: play		Nominal	Nominal	Nominal	Nominal	Nominal	1	sunny	hot	high	FALSE	no	2	sunny	hot	high	TRUE	no	3	overcast	hot	high	FALSE	yes	4	rainy	mild	high	FALSE	yes	5	rainy	cool	normal	FALSE	yes	6	rainy	cool	normal	TRUE	no	7	overcast	cool	normal	TRUE	yes	8	sunny	mild	high	FALSE	no	9	sunny	cool	normal	FALSE	yes	10	rainy	mild	normal	FALSE	yes	11	sunny	mild	normal	TRUE	yes	12	overcast	mild	high	TRUE	yes	13	overcast	hot	normal	FALSE	yes	14	rainy	mild	high	TRUE	no		
	1: outlook	2: temperature	3: humidity	4: windy	5: play																																																																																																
	Nominal	Nominal	Nominal	Nominal	Nominal																																																																																																
1	sunny	hot	high	FALSE	no																																																																																																
2	sunny	hot	high	TRUE	no																																																																																																
3	overcast	hot	high	FALSE	yes																																																																																																
4	rainy	mild	high	FALSE	yes																																																																																																
5	rainy	cool	normal	FALSE	yes																																																																																																
6	rainy	cool	normal	TRUE	no																																																																																																
7	overcast	cool	normal	TRUE	yes																																																																																																
8	sunny	mild	high	FALSE	no																																																																																																
9	sunny	cool	normal	FALSE	yes																																																																																																
10	rainy	mild	normal	FALSE	yes																																																																																																
11	sunny	mild	normal	TRUE	yes																																																																																																
12	overcast	mild	high	TRUE	yes																																																																																																
13	overcast	hot	normal	FALSE	yes																																																																																																
14	rainy	mild	high	TRUE	no																																																																																																
	OR																																																																																																				
18	Explain Naïve-Bayers classifiers and predict instance to the class (weather data set) if outlook=sunny, temp=mild, humidity=normal, windy=true, play =?	[10M]	4	L2																																																																																																	
19	Explain the outlier detection methods.	[10M]	5	L2																																																																																																	
	OR																																																																																																				
20	Discuss key issues involved in hierarchical and K-Means clustering	[10M]	5	L2																																																																																																	