2022/ODD/18/24/MCSC-903/MS-303/336

UG/PG ODD SEMESTER EXAMINATION, 2022

COMPUTER SCIENCE

9th / 3rd Semester

Course No. : MSCC-903/MS-303

(Data Mining & Knowledge Discovery)

Full Marks : 70

Pass Marks : 28

Time : 3 hours

The figures in the margin indicate full marks for the questions

(Answer any five questions, taking one from each unit)

UNIT-I

1. (a) Does a Data Warehouse require transaction processing, concurrency control and recovery? Explain why. 3

   (b) Why encoding of data is often found to be inconsistent? How data are made consistent while implementing Data Warehouse? 3

   (c) Using a representative data cube define 'SLICING' and 'DICING' operations and give one example for each. 2+2=4

   (d) State the limitation of virtual Data Warehouse. Are there any advantage(s) of with virtual Data Warehouse? State. 2+2=4

2. (a) Distinguish between 'SNOW FLAKE SCHEMA" and 'FACT CONSTELLATION SCHEMA' for Data Warehouse design. 4

(Turn Over)

(b) (i) Consider the following data (SALES DATA) and construct a Data Warehouse using multi-dimensional data model. 6

(ii) Perform SLICING and DICING operations on this Data Warehouse by writing your own query. 2+2=4

SALES DATA
LOCATI\ON

ITEM TYPE		Year	Kolkata	Mumbai	Chennai	New Delhi	Hyderabad
ITEM TYPE	TV	91	1977	2411	2341	3261	2952
		92	2099	2870	2365	3555	3061
		93	2237	3232	2455	3626	3223
		94	2354	3594	2771	3777	3571
		95	2078	3810	2691	3400	3622
	AUDIO	91	3120	3235	3025	4120	3621
		92	2420	2943	3122	4020	3502
		93	2890	2945	2942	4110	3142
		94	1825	2715	2767	3912	3612
		95	1729	3923	2892	3990	3341
	COPUTER	91	2345	4311	4100	3995	3772
		92	3123	4132	3920	3925	3612
		93	3622	4626	3712	3840	3662
		94	2842	4011	3850	4028	3419
		95	3242	3800	3760	4098	3526

UNIT-II

3. (a) What are the obtained as the output of Data Mining process? What are their characteristics? 4

(b) State what can be obtained as a result of mining patient data and their medical diagnosis research? 2

005	45	60
006	30	40
007	50	60
008	60	80

UNIT-V

9. Consider the following dataset and apply the ID3 algorithm to classify the data using information gain technique based on entropy by the method of decision tree learning. 14

					CLASS LEBEL
Date	Outlook	Temperature	Humidity	Windy	Play Tennis
01	Sunny	Hot	High	No	No
02	Sunny	Hot	High	Yes	No
03	Overcast	Hot	High	No	Yes
04	Rain	Mild	High	No	Yes
05	Rain	Cool	Normal	No	Yes
06	Rain	Cool	Normal	Yes	No
07.	Overcast	Cool	Normal	Yes	Yes
08	Sunny	Mild	High	No	No
09	Sunny	Cool	Normal	No	Yes
10	Rain	Mild	Normal	No	Yes
11	Sunny	Mild	Normal	Yes	Yes
12	Overcast	Mild	High	Yes	Yes
13	Overcast	Hot	Normal	No	Yes
14	Rain	Mild	High	Yes	No.

10. (a) What is web structure mining? Explain the working of the Page Rank algorithm. 2+8=10

(b) Explain the underlying data structure for Page Rank algorithm. 4

****

(c) Distinguish between DBMS and Data Mining. 4

(d) Define association rules? How the interestingness of associaion rules are measured? give examples. 4

4. (a) Give examples of discovery dirver and verification driver taks of data mining? Why are these called so? 3

(b) What are the neural networks? State the short comings of application of neural networks for classification of data. 4

(c) What is spatial data mining? Explain. 4

(d) Distinguish between clustering and classification techniques for data mining. 3

UNIT-III

5. Consider the following transaction database and apply the apriori algorithm to discover at least 5 association rules with minimum support=20% and minimum confidence= 50%. Show all the steps of iteration of the a priori algorithm for all the database scans. 14

TID	Item				
001	a	b	c		
002		b	c	d	e
003	a		c		e
004		b	c		e
005	a		c		e
006	b			d	e
007	a		c	d	
008		b	c	d	
009			c	d	e
010	a	b	c	d	e

6. How the Pincer-Search Algorithm improve upon the a priori algorithm? Explain the working of the Pincer-Search algorithm by using the dataset given in Q5 above. Find all the outputs for the above dataset. 14

UNIT-IV

7. (a) Explain the principle of FP Tree Growth algorithm. How is it different from the a priori algorithm? 4+2=6

(b) What is agglomerative clustering? How it is different from divisive clustering? 2+1=3

(c) Why conventional clustering methods which compute only numeric data are not suitable for data mining? 2

(d) State at least three major issues concerning the efficiency of the clustering algorithm for data mining. 3

8. Write the k-means clustering algroithm and by applying k-means clustering algorithm cluster the following data into 02 clusters. (Do at least 4 iterations) 14

RIP	Age	Salary (Rs.)
001	25	20
002	35	40
003	40	38
004	50	75