8. a) State the principles of partitioning algorithm for clustering. 2

   b) Explain the working of the PAM algorithm for clustering? 12

UNIT - V

9. What is categorical clustering? Explain the Robust Hierarctical-clustering with links (ROCK) algorithm. 2+12=14

10. a) What is Page Rank? Explain a page rank algorithm with one example. 8

    b) Explain the features of unstructured text. 6

PG / INTEGRATED (CBCS) ODD SEMESTER EXAMINATION, 2021
Held in April 2022

COMPUTER SCIENCE
9th Semester / 3rd Semester

COURSE NO. MCSCC - 903 / MS - 303
(Data Mining and 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) What is dimensional hierarchy? How dimensional modelling is useful for data warehouse? 2+2=4

   b) Design a data warehouse by using multidimensinal data model with the following data

		Students' Admission (in thousands)									
		Year	School			College (UG)			University (PG)		
			Science	Arts	Commerce	Non Engg.	Engg.	Medical	Non Engg.	Engg. PG	Medical PG
GENDER	Male	2011	901	5032	3021	6053	1755	232	2122	312	75
		2012	2215	5821	3221	7031	1821	242	2015	342	77
		2013	2421	5622	3011	6892	1890	253	2211	363	81
		2014	2344	5821	2942	7112	1921	251	2345	292	79
		2015	2400	6000	2802	8012	2012	232	2401	282	76

GENDER	Female	2011	1621	6121	1321	6211	1521	192	2201	186	46
		2012	1811	5911	1299	6012	1611	191	2045	176	48
		2013	1842	6220	1175	6221	1605	185	2126	182	45
		2014	1711	6120	1221	5900	1700	173	2542	180	42
		2015	1706	5942	1311	5741	1688	170	2412	172	41

Base on the data warehouse designed using the above data, write one query each SLICE, DICE and ROLLUP operations and show the output by solving the queries. 4+2+2+2=10

2. a) Why a data warehouse is called an integrated and non volatile collection of data? 3

   b) What is lattice of cuboids? Explain. 3

   c) What modifications are made in the SNOWFLAKE schema as compared to the STAR schema? Explain. 3

   d) Is implementing a data warehouse prerequisite for performing data mining? Justify your answer. 2

   e) Distinguish between ROLAP and MOLAP. 3

UNIT - II

3. a) Explain the need of data mining? Is it possible to perform data mining by using SQL? 3+2=5

   b) Is it possible to perform data mining using neural network techniques? Explain. 3

   c) Distinguish between supervised and unsupervised learning techniques for data mining. 3

   d) What are genetic algorithm? How are these useful for data mining? 3

4. a) What do you mean by rough sets? Explain. 3

   b) What is clustering? Explain with an example. 3

   c) Define frequent episodes and its types. 4

   d) Explain major challenges of data mining. 4

UNIT - III

5. a) What is lattice of subsets? Construct a lattice of subsets for the set A = {a, b, c, d, e} 4

   b) Explain the working of the apriori algorithm for the discovery of association rules with the help of an example. 10

6. a) What is Maximal Frequent Sets (MFS) and Border Sets? How such sets of item can influence the algorithm for discovery of frequent pattern. Explain. 4

   b) Explain the principle and working of the partition algorithm for discovery of association rules. 10

UNIT - IV

7. Explain the principle and working of the Pincer-search algorithm for frequent item set mining. 14