

IV YEAR B.TECH. EXAMINATION DEC 2022**150712: Data Mining & Warehousing**

Time: 3 Hours Maximum Marks: 70 Minimum Passing Marks: 22

Note:		1. Answer all five questions. All questions carry equal marks. 2. In each question part a, b, c is compulsory and part d has internal choice. Out of which part a & b carry 2 marks, part c carry 3 marks and part d carry 7 marks 3. All Parts of each question are to be attempted at one place 4. Assume suitable value for missing data, if any		
Question No.		Mark s	Course Outcomes	Bloom's Level
1. (a)	How is Data warehouse different from a database?	02	C01	L1
(b)	What is the difference between classification & prediction?	02	C02	L1
(c)	Explain data mining task with an example?	03	C02	L2
(d)	Briefly explain the major issues in data mining?	07	C03	L2
(e)	Explain the following i. Association Analysis ii. Outlier Analysis iii. Spatial databases	07	C01,C02	L3
2. (a)	What is a Data warehouse?	02	C02	L2
(b)	What is Data Mart?	02	C02	L3
(c)	Compare operational database system & data warehouses?	03	C02	L2
(d)	Explain briefly data cube technology?	07	C03,C04	L3
(e)	Briefly compare the following concepts: Snowflake schema, fact constellation, starburst query model.	07	C04	L4
3. (a)	What do you mean by data cleaning?	02	C03	L3
(b)	What is numerosity reduction?	02	C03	L2
(c)	What is Entity identification problem?	03	C05	L3
(d)	Use these methods to normalize the following group of data: 200,300, 400, 600,1000 (a) Min -max normalization by setting min=0 and max=1 (b) Z-score normalization (c) Normalization by decimal scaling	07	C03	L4
(e)	Explain the following: (a) Stepwise forward selection (b) Stepwise backward elimination (c) A combination of forward selection and backward elimination	07	C02	L3
4. (a)	How do you measure the quality of an association rule?	02	C05	L4
(b)	Discuss Apriori property?	02	C05	L3
(c)	How to generate association rules, discuss?	03	C04	L4
(d)	Explain how to improve the Efficiency of Apriori Algorithm?	07	C05	L5
(e)	Discuss multidimensional association rules and constraint based association rule mining?	07	C06	L1

5.	(a) What are the issues in classification?	02	C06	L1
	(b) What do you mean by CART?	02	C06	L2
	(c) Explain similarity and distance measures?	03	C01,C04	L1
	(d) Explain (a). K-means clustering (b). DBSCAN algorithm	07	C06	L4
	(e) Compare some of the important clustering algorithms?	07	C06	L6

