BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: MCA **SEMESTER : II**
BRANCH: MCA **SESSION : SP/2023**

SUBJECT: CA441 DATA MINING TECHNIQUES

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Describe four challenges of mining regarding data mining methodology.	[5]	1	2
Q.1(b)	Differentiate between (i) Classification & regression and (ii) Characterization & Clustering.	[5]	1	2
Q.2(a)	Compare Snowflakes and Fact Constellation schema for multidimensional data model.	[5]	2	2
Q.2(b)	Explain Data Cube: A multidimensional Data Model with a neat diagram.	[5]	2	3
Q.3(a)	What is an association rule? Define Support, Confidence & Large itemset.	[5]	3	2
Q.3(b)	Consider the dataset in Table 1.Using the Apriori algorithm, find all frequent itemsets and generate association rules with a minimum support of three.	[5]	3	3

Table 1

Transaction ID	Items
T1	a, b, e
T2	b, c, d, e
T3	a, b, c
T4	a, b, c, d
T5	a, b, c

Q.4(a)	What are the classification issues? Use an example to demonstrate.	[5]	4	2
Q.4(b)	Explain decision tree classification using an example. Use entropy and information gain.	[5]	4	3
Q.5(a)	What is outlier analysis? What is the different classification of clustering methods?	[5]	5	2
Q.5(b)	Using the K-means algorithm and the Euclidean distance formula, Find two clusters in the following data: D={(2,4), (3,5), (4,3),(4,5), (6,7), (5,6), (7,5)}	[5]	5	3

:::::01/05/2023 E:::::