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

CLASS: MCA **SEMESTER : II**
BRANCH: MCA **SESSION : SP/2025**

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)	What is data mining? Enlist the steps involved in data mining when viewed as a process of knowledge discovery.	[5]	1	2
Q.1(b)	Explain different steps for data processing in Data Mining.	[5]	1	4
Q.2(a)	With suitable examples give a comparative analysis between Star schema, Snowflake schema and fact constellations.	[5]	2	4
Q.2(b)	Describe the components of three tier data warehouse architecture.	[5]	2	2
Q.3(a)	Illustrate the working of Apriori algorithm with the following Transactional Data. Minimum Support Count = 2	[5]	3	4

TID	List of item IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Q.3(b)	Explain the concept of multilevel association mining. Justify with examples where this can be applied.	[5]	3	5
Q.4(a)	Briefly outline the major steps of decision tree classification. Illustrate with an example.	[5]	4	4
Q.4(b)	Why is Naive Bayesian classification called "naive"? Briefly outline the major ideas of Naive Bayesian classification.	[5]	4	3
Q.5(a)	Apply the K-Means Algorithm to create two clusters (K=2) for the following dataset: $\{A(1,1), B(2,1), C(4,2), D(5,3), E(4,4)\}$. Consider *A* and *B* as the initial "*means*". Compute the Sum squared error at each iteration.	[5]	5	3
Q.5(b)	Describe any one: (i) DBSCAN algorithm (ii) Outlier detection	[5]	5	2

::::::02/05/2025::::::E