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

CLASS:MTech/Pre-PhD **SEMESTER : I**
BRANCH: Computer Science & Engineering **SESSION : MO/2025**

SUBJECT: CS539 DATA MINING AND DATA WAREHOUSING

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)	Explain the major issues and challenges in data mining with suitable examples.	[5]	1	1
Q.1(b)	A bank applies data mining to its transaction logs and obtains: Normal transaction patterns: 900 Suspicious patterns: 60 Alert-triggering patterns: 30 Confirmed fraud patterns: 10 (i) Compute the percentage distribution of each pattern type. (ii) Comment on the effectiveness of the bank's fraud detection system based on the distribution.	[5]	1	4
Q.2(a)	A public opinion poll surveyed a simple random sample of 1000 voters. Respondents were classified by gender (male or female) and by voting preference (Republican, Democrat, or Independent). Results are shown in the contingency table below.	[5]	2	3

GENDER	VOTING PREFERENCES		
	REPUBLIC	DEMOCRAT	INDEPENDENT
MALE	200	150	50
FEMALE	250	300	50

Obtain Correlation among the attributes using the Chi-Square method,
(Given for 2 degrees of freedom Chi-Square value to reject the hypothesis at the 0.005 significance level is 10.597)

Q.2(b)	Briefly outline how to compute the dissimilarity between objects described by the following: (a) Nominal attributes (b) Asymmetric binary attributes (c) Numeric attributes (d) Term-frequency vector	[5]	2	2
Q.3(a)	A small dataset has 4 observations with 2 attributes: X_1(Annual Income in ₹10,000s) and X_2(Monthly Spending in ₹1,000s).	[5]	3	4

Observation	X_1	X_2
1	4	5
2	6	7
3	8	9
4	10	11

Assume the data is to be analysed using **Principal Component Analysis (PCA)**.

Q.3(b)	Briefly compare the following concepts. You may use an example to explain your point(s). i) Data cleaning ii) Data transformation	[5]	3	2

PTO

Q.4(a)	Explain star, snowflake, and fact constellation schemas with diagrams for a Product-Based Company.	[5]	4	5
Q.4(b)	A 3D sales cube has: Product: 20 Region: 16 Time: 12 months Initially, it's stored at the month level and is 30% dense. Compute the base cube cell count (month-level). If data is rolled up from Month → Quarter (4 quarters per year), what is the new cube size (number of cells at quarter granularity)? Assuming density remains 30%, how many non-empty cells are there at: (a) Month level (b) Quarter level	[5]	4	5
Q.5(a)	Dataset (10 transactions) T1: milk, bread, egg T2: bread, butter T3: milk, bread, butter T4: bread, egg T5: milk, egg, cheese T6: bread, milk, butter, egg T7: butter, cheese T8: milk, bread, cheese T9: milk, butter T10: bread, cheese (a) Using minimum support = 30% (min count = 3), list frequent items and their counts. (b) Find the frequent itemsets in FP-growth.	[5]	5	4
Q.5(b)	Generate rules from 5(a), compare with support, confidence, lift, and all_confidence.	[5]	5	5

:::::25/11/2025:::::E