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

CLASS: BBA **SEMESTER : V**
BRANCH: BBA **SESSION : MO/2025**

SUBJECT: MN435 DATA MINING

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 different Data Mining tasks. How do predictive and descriptive data mining tasks differ?	[5]	1	1
Q.1(b)	What is data cleaning? Explain its importance and common techniques used for handling noisy or missing data.	[5]	1	2
Q.2(a)	Find the Hamming Distance between the following pairs of binary strings: a) X = 1011101 and Y = 1001001 b) X = 2173896 and Y = 2233796	[5]	2	3
Q.2(b)	Calculate the Cosine Similarity between the following two document vectors: x=(4,0,3,0,2,0,0,1,0) y=(2,0,0,0,1,0,0,0,3)	[5]	2	3
Q.3(a)	Define Association Analysis. How does it help in market basket analysis? Explain the importance of support, confidence in evaluating association rules.	[5]	3	1,2
Q.3(b)	What is the Apriori algorithm? Explain how it is used for association rule mining in data mining.	[5]	3	1,2
Q.4(a)	Explain Decision Tree? Describe its key components- root node, internal nodes, leaf nodes, and branches with the help of an example.	[5]	4	1,2
Q.4(b)	A marketing company developed a model to predict whether customers will buy a new product or not. After testing the model on 100 customers, the following results were obtained: 50 customers who actually bought the product were correctly predicted as buyers. 10 customers who actually bought the product were wrongly predicted as non-buyers. 5 customers who did not buy the product were wrongly predicted as buyers. 35 customers who did not buy the product were correctly predicted as non-buyers. Based on this information, draw the confusion matrix and calculate the following performance measures of the model: Accuracy, Error Rate, Precision, Recall, F1 Score	[5]	4	3
Q.5(a)	A digital marketing firm wants to group its customers based on their online buying patterns. To analyse the customer groups, the firm applies Hierarchical Clustering. (a) Explain the difference between Agglomerative and Divisive hierarchical clustering methods in this context. (b) Which method would be more suitable for this firm and why?	[5]	5	2,4
Q.5(b)	Write short notes on the following: (a) Web Mining (b) Spatial Data Mining	[5]	5	1

::::::28/11/2025::::::M