Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 07/10/2026 **Time:** 09:30 AM - 11:00 AM

Course Code: CSE3512 **Course Name:** Data Mining

Semester: Fifth Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2	CO3
Marks	56	76	48

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	Explain any 4 data mining tasks.	2	CO1	BT3
2	Explain any 2 data mining techniques with examples	2	CO1	BT2
3	Perform data smoothing by bin boundaries on 3 equi-width bins. Data: [24,27,29,16,17,31,33,29,36,37,35,44]	2	CO2	BT3
4	Explain the two parameters of Association Rule Mining.	2	CO3	BT2
5	In a market basket dataset of 10 transactions, an itemset `{Bread, Milk}` occurs together in 4 transactions. Calculate the support of the itemset.	2	CO3	BT3

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	Explain the KDD process in Data Mining with appropriate diagram	10	CO1	BT2
7	Explain the stages of data mining process with suitable diagram/ flowchart.	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	Consider the following dataset with two features **X** and **Y**: **Object X Y** A 2 3 B 4 5 C 6 4 D 8 7 E 5 6 For the data point P = (6, 5): a) Calculate the mean vector of the two features. b) Calculate the covariance matrix. c) Calculate the Mahalanobis distance between P and the mean vector.	10	CO2	BT3

9	Consider two data points with features x and y: point x y p1 0 2 p2 2 0 p3 3 1 p4 5 1 Calculate the Minkowski distance between the points Q for the following values of r . 1.r=1 Manhattan Distance 2.r=2 Euclidean Distance 3.r=∞ Chebyshev Distance	10	CO2	BT3

PART D: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
10	Consider the following transaction database of 5 supermarket transactions. Given a Minimum Support Threshold = 40% (Support Count>=2) and a Minimum Confidence Threshold = 60%. TID Items 1 Bread, Milk 2 Bread, Diaper, Beer 3 Milk, Diaper, Beer 4 Bread, Milk, Diaper 5 Bread, Milk, Coke 1. Generate all frequent itemsets using the Apriori algorithm. 2. Extract all strong association rules and calculate their Support and Confidence.	10	CO3	BT3
11	Consider the above transaction database of 5 supermarket transactions from Q.N.:10 a) Calculate the support count and support for the itemset {Bread, Milk}. b) Calculate the confidence for the association rule Bread → Milk. c)Given a minimum support threshold of 60%, determine whether the following itemsets are frequent itemsets. Justify your answers based on their support values: 1.{Bread, Milk} 2.{Bread, Butter}	10	CO3	BT3

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	Consider the following two data objects A and B **Attribute Object A Object B** Age 20 25 Income 30 40 Purchases 4 7 a)Calculate the Euclidean distance between A and B. b) Calculate the Cosine similarity between the two objects by treating their attribute values as vectors. Consider the following sets of products purchased by two customers: Customer A: {Bread, Milk, Eggs, Butter} Customer B: {Bread, Milk, Cheese, Butter} c) Calculate the Jaccard similarity between Customer A and Customer B. d) State what a higher value of Cosine and Jaccard similarity indicates.	10	CO2	BT3

13	a)The values of the attribute Age for five students are 20, 25, 30, 35, 40 .Using Min-Max Normalization, transform the value 30 to the range [0, 1]. b)The mean and standard deviation of the marks obtained by a group of students are 60 and 10, respectively. Using Z-score standardization, determine the standardized value of a student who scored 80 marks.	10	CO1	BT2