

- Note:** (a) All questions are compulsory.
 (b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems
 (c) Calculator is allowed.

Q. No.	Question	CO	Marks																					
Q1.	a. Analyze the key characteristics of a data warehouse and evaluate how it supports the KDD process in extracting valuable insights for decision-making.	CO1	[3]																					
	b. Consider the data: 25, 30, 45, 50, 60. Apply Min-Max normalization method on this data to scale the values between -1 and 1.		[2]																					
Q2.	a. Consider two text documents represented as term frequency vectors in a 6-dimensional space: Doc1 = (2,3,0,5,7,1), Doc2 = (4,1,2,6,3,2) Compute the cosine similarity of these two documents. Also interpret the results.	CO3	[2]																					
	b. Consider the daily step count (in thousands) of two fitness enthusiasts over five days: Person A = (7, 10, 12, 9, 11), Person B = (8, 9, 14, 10, 12) Compute the Manhattan distance between their step counts. Interpret the results in terms of the similarity of their physical activity levels.		[2]																					
Q3.	a. How do you determine the value of k in the k-NN algorithm? What are the drawbacks of choosing a value of k that is too small?	CO4	[2]																					
	b. Given the dataset below, apply ID3 algorithm to determine the root node of the decision tree.		[4]																					
	<table border="1"> <thead> <tr> <th>Age</th> <th>Income</th> <th>Buy Product?</th> </tr> </thead> <tbody> <tr> <td>Young</td> <td>High</td> <td>No</td> </tr> <tr> <td>Young</td> <td>Low</td> <td>Yes</td> </tr> <tr> <td>Middle</td> <td>High</td> <td>Yes</td> </tr> <tr> <td>Middle</td> <td>Low</td> <td>Yes</td> </tr> <tr> <td>Old</td> <td>High</td> <td>Yes</td> </tr> <tr> <td>Old</td> <td>Low</td> <td>No</td> </tr> </tbody> </table>	Age	Income	Buy Product?	Young	High	No	Young	Low	Yes	Middle	High	Yes	Middle	Low	Yes	Old	High	Yes	Old	Low	No		
Age	Income	Buy Product?																						
Young	High	No																						
Young	Low	Yes																						
Middle	High	Yes																						
Middle	Low	Yes																						
Old	High	Yes																						
Old	Low	No																						