



Daffodil International University
Faculty of Science & Information Technology
Department of Computer Science & Engineering
Final Semester Examination, Fall 2024
Course Code: CSE325 , Course Title: Data Mining & Machine Learning
Level: 3 Term: 2 Batch: 61

Time: 2:00 Hrs

Marks: 40

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	a)	Analyze the following transaction dataset to derive all frequent itemsets with a minimum support & Confidence threshold of 50%. Justify your approach by evaluating the intermediate steps and explaining the computational challenges encountered.	[Marks:10]	CO3																						
		<table><tr><th>Transaction ID</th><th>Items Purchased</th></tr><tr><td>T1</td><td>Bread, Butter, Milk, Eggs</td></tr><tr><td>T2</td><td>Bread, Butter, Jam</td></tr><tr><td>T3</td><td>Milk, Bread, Butter, Eggs</td></tr><tr><td>T4</td><td>Butter, Milk, Eggs</td></tr><tr><td>T5</td><td>Bread, Butter, Eggs, Jam</td></tr><tr><td>T6</td><td>Bread, Milk, Eggs</td></tr><tr><td>T7</td><td>Bread, Butter, Milk, Jam</td></tr><tr><td>T8</td><td>Butter, Jam</td></tr><tr><td>T9</td><td>Bread, Butter, Milk, Eggs</td></tr><tr><td>T10</td><td>Butter, Milk, Jam</td></tr></table>	Transaction ID	Items Purchased	T1	Bread, Butter, Milk, Eggs	T2	Bread, Butter, Jam	T3	Milk, Bread, Butter, Eggs	T4	Butter, Milk, Eggs	T5	Bread, Butter, Eggs, Jam	T6	Bread, Milk, Eggs	T7	Bread, Butter, Milk, Jam	T8	Butter, Jam	T9	Bread, Butter, Milk, Eggs	T10	Butter, Milk, Jam		
Transaction ID	Items Purchased																									
T1	Bread, Butter, Milk, Eggs																									
T2	Bread, Butter, Jam																									
T3	Milk, Bread, Butter, Eggs																									
T4	Butter, Milk, Eggs																									
T5	Bread, Butter, Eggs, Jam																									
T6	Bread, Milk, Eggs																									
T7	Bread, Butter, Milk, Jam																									
T8	Butter, Jam																									
T9	Bread, Butter, Milk, Eggs																									
T10	Butter, Milk, Jam																									
2.	a)	Explain the concept of cross-validation and discuss how it contributes to improving a model's generalization ability on new data.	[Marks:3]	CO2																						
	b)	Demonstrate the challenges of using regular K-Fold Cross-Validation with a highly imbalanced dataset. Show how Stratified K-Fold Cross-Validation can be applied to resolve these issues effectively.	[Marks:7]																							
3.	a)	Define Hierarchical Agglomerative Clustering (HAC) and identify any two methods used to measure the distance between clusters in HAC.	[Marks:2]	CO3																						
	b)	Construct a Dendrogram for the dataset below using HAC with single-linkage. Data-points are A1(1,3),A2(2,4),A3(3,5),A4(7,8),A5(8,9),B1(1,2),B2(4,6).	[Marks:8]																							

4.	a)	Analyze Forward Propagation on the following Neural Network given below and calculate the loss. The actual output of $y_8=0.01$ and $y_9=0.99$ and the neurons are using sigmoid activation function.	[Marks:10]	CO4
----	----	---	------------	-----

