

**Model Question Paper**  
**Second Semester MCA Degree Examination, 2025-26**

**MACHINE LEARNING**

**Time: 3 Hours**

**Max. Marks: 100**

*Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.  
2. M: Marks, L: RBT (Revised Bloom's Taxonomy) level, C: Course outcomes.*

		Module -1	M	L	C																																			
Q1	a.	<p>An online retail store wants to analyze customer purchasing patterns using association rule mining. It is observed that customers who buy Tea may also buy Sugar. The transaction data is given below:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Transaction ID</th> <th>Items Purchased</th> </tr> </thead> <tbody> <tr><td>T1</td><td>Tea, Sugar</td></tr> <tr><td>T2</td><td>Tea</td></tr> <tr><td>T3</td><td>Tea, Sugar</td></tr> <tr><td>T4</td><td>Sugar</td></tr> <tr><td>T5</td><td>Tea, Sugar</td></tr> <tr><td>T6</td><td>Tea</td></tr> </tbody> </table> <p>Calculate the support, confidence, and lift for the rule Tea → Sugar. Based on these values, determine whether this rule indicates a strong association. Write a brief with your conclusions.</p>	Transaction ID	Items Purchased	T1	Tea, Sugar	T2	Tea	T3	Tea, Sugar	T4	Sugar	T5	Tea, Sugar	T6	Tea	10	L3	CO1																					
	Transaction ID	Items Purchased																																						
T1	Tea, Sugar																																							
T2	Tea																																							
T3	Tea, Sugar																																							
T4	Sugar																																							
T5	Tea, Sugar																																							
T6	Tea																																							
b.	<p>A Smart Fitness Monitoring System collects health-related data from users as shown below:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>User</th> <th>Age</th> <th>Heart Rate</th> <th>Blood Pressure</th> <th>Smoker</th> <th>Risk</th> </tr> </thead> <tbody> <tr><td>1</td><td>35</td><td>78</td><td>118</td><td>Yes</td><td>Low</td></tr> <tr><td>2</td><td>42</td><td>—</td><td>125</td><td>yes</td><td>Medium</td></tr> <tr><td>3</td><td>180</td><td>250</td><td>400</td><td>No</td><td>High</td></tr> <tr><td>4</td><td>29</td><td>70</td><td>110</td><td>NO</td><td>Low</td></tr> <tr><td>5</td><td>35</td><td>78</td><td>118</td><td>Yes</td><td>Low</td></tr> </tbody> </table> <p>Identify and correct the data quality issues using appropriate data cleaning techniques. Justify your approach and provide the final cleaned dataset.</p>	User	Age	Heart Rate	Blood Pressure	Smoker	Risk	1	35	78	118	Yes	Low	2	42	—	125	yes	Medium	3	180	250	400	No	High	4	29	70	110	NO	Low	5	35	78	118	Yes	Low	10	L3	CO1
User	Age	Heart Rate	Blood Pressure	Smoker	Risk																																			
1	35	78	118	Yes	Low																																			
2	42	—	125	yes	Medium																																			
3	180	250	400	No	High																																			
4	29	70	110	NO	Low																																			
5	35	78	118	Yes	Low																																			
<b>OR</b>																																								
	a.	<p>The dataset represents employee attrition prediction in a company based on various features:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Employee</th> <th>Salary</th> <th>Experience (Years)</th> <th>Performance Score</th> <th>Attrition</th> </tr> </thead> <tbody> <tr><td>1</td><td>25000</td><td>2</td><td>3.5</td><td>No</td></tr> <tr><td>2</td><td>40000</td><td>5</td><td>4.2</td><td>Yes</td></tr> </tbody> </table>	Employee	Salary	Experience (Years)	Performance Score	Attrition	1	25000	2	3.5	No	2	40000	5	4.2	Yes	10	L3	CO1																				
Employee	Salary	Experience (Years)	Performance Score	Attrition																																				
1	25000	2	3.5	No																																				
2	40000	5	4.2	Yes																																				

Q2		3	32000	3	3.8	—					
		4	60000	8	4.5	No					
		5	25000	2	3.5	No					
		Apply different feature selection techniques to identify important features. Compare the results and justify which features are most relevant.									
b.	A company has recorded the monthly salary (₹) and years of experience of its employees as shown below:								10	L3	CO1
		Employee		Salary (₹)	Experience (Years)						
		E1		18000	1						
		E2		22000	2						
		E3		26000	3						
		E4		35000	5						
		E5		150000	12						
	Apply Normalization, Standardization, and Robust Scaling on the Salary feature and determine the most suitable method with justification.										

### Module- 2

Q3	a.	In the context of Learning Multiple Classes, construct and compare the One-vs-All (OvA) and One-vs-One (OvO) approaches used for multi-class classification. You should include suitable examples to explain how each method works and discuss their advantages and limitations.								10	L3	CO2
	b.	A medical researcher wants to study the relationship between Age (in years) and Blood Glucose Level (mg/dL). The following data is collected:								10	L3	CO2
		Age (X)		Glucose Level (Y)								
		25		90								
		30		95								
		35		100								
		40		110								
		45		120								
	Using the method of simple linear regression, find the regression equation of the form $Y = a + bX$ . Predict the blood glucose level for a person aged 50 years.											

### OR

Q4	a.	Marie is getting married tomorrow at an outdoor ceremony in the desert, where it historically rains only 5 days a year. However, the weather forecast predicts rain for tomorrow. Past data shows that the forecast correctly predicts rain 90% of the time when it actually rains, and incorrectly predicts rain 10% of the time when it does not rain. Apply Bayes' Theorem to determine the probability that it will actually rain tomorrow given that rain has been forecasted.								10	L3	CO2
	b.	Explain Maximum Likelihood Estimation (MLE) and apply it to estimate parameters in Bernoulli, Multinomial, and Gaussian (Normal) density models with suitable examples.								10	L3	CO2

### Module - 3

Q5	a.	Compare k-Means Clustering and Hierarchical Clustering with respect to cluster formation, working procedure, computational complexity, advantages, and limitations. Analyze how both techniques are utilized for grouping similar data objects in unsupervised learning applications.								10	L3	CO3
----	----	---	--	--	--	--	--	--	--	----	----	-----

	b.	Differentiate Histogram Estimator, Kernel Estimator, and k-Nearest Neighbor Estimator based on their working principles, density estimation process, parameter selection, and applications. Examine the role of Nonparametric Density Estimation techniques in data analysis.			
<b>OR</b>					
Q6	a.	Construct Univariate Trees and Classification Trees by selecting appropriate splitting attributes and modeling the decision-making process for classification problems.	10	L3	CO3
	b.	Build an Unsupervised Learning model by planning, organizing, and utilizing clustering and density estimation techniques to solve real-world data analysis problems.	10	L3	CO3
<b>Module - 4</b>					
Q7	a.	Illustrate the role of Neural Networks as a paradigm for parallel processing by relating the structure of artificial neural networks with the functioning of the human brain. Examine the importance of neural networks in machine learning applications.	10	L3	CO4
	b.	Construct the architecture of a Perceptron and illustrate the procedure involved in training a Perceptron model. Illustrate the significance of learning rules in pattern classification problems.	10	L3	CO4
<b>OR</b>					
Q8	a.	Develop a Multilayer Perceptron model and illustrate the working of the Backpropagation Algorithm for weight updation and error minimization in neural network training.	10	L3	CO4
	b.	Differentiate Nonlinear Regression, Two-Class Discrimination, and Multiclass Discrimination using Multilayer Perceptrons. Examine how neural network models are utilized for solving complex classification and prediction problems.	10	L3	CO4
<b>Module - 5</b>					
Q9	a.	Use the elements of Reinforcement Learning and examine the interaction between agent, environment, states, actions, and rewards in the learning process.	10	L3	CO5
	b.	Develop a Model-Based Reinforcement Learning approach by analyzing the working procedure of Value Iteration for obtaining optimal policies. Examine the significance of model representation and iterative value updates in sequential decision-making problems.	10	L3	CO5
<b>OR</b>					
Q10	a.	Apply the concepts of deterministic and nondeterministic rewards and actions to model a reinforcement learning environment. Illustrate with an example how an agent selects actions and learns policies under both deterministic and uncertain conditions.	10	L3	CO5
	b.	Apply Temporal Difference Learning, Eligibility Traces, Exploration Strategies, and Generalization techniques to design a reinforcement learning approach for an agent operating in a complex environment. Explain how these techniques improve learning efficiency and help balance exploration and exploitation.	10	L3	CO5

\*\*\*\*\*