



Daffodil International University

Faculty of Science & Information Technology

Department of Computer Science & Engineering

Midterm Examination, Fall 2026

Course Code: CSE431, Course Title: Machine Learning

Level: 4 Term: 2, Batch: 65

Time: 01:30 Hrs

Marks: 25

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.]

A manufacturing company is developing a Machine Learning system to predict motor failure (Yes/No) within the next hour. The system makes predictions at 9:00 AM using available machine data. Historical failure records are available for training, and the company prefers an interpretable model that provides failure probabilities.

During dataset preparation, the engineering team observes the following:

1. Sensor Records: The recorded motor temperatures include 60°C, 62°C, 64°C, and 120°C, along with several missing readings. All the values are verified as a genuine measurement. The company needs to replace missing readings using a representative value that is not strongly influenced by unusually high temperatures. Due to network retransmission, some sensor records also appear multiple times with identical Motor IDs, timestamps, and measurements.

2. Motor Information: The factory operates three motor types: A, B, and C. These represent different motor designs without any natural ranking. An engineer has assigned the values 1, 2, and 3 to these types for model training.

3. Operating Conditions: Two motors have the following readings at 9:00 AM. The engineering team has also calculated a new feature, Temperature Difference = Motor Temperature – Room Temperature, to capture the difference between motor and surrounding temperatures.

Feature	Motor 1	Motor 2
Motor Temperature	80°C	80°C
Room Temperature	20°C	40°C

The database also contains a field named RepairConfirmed, which becomes available only after 10:00 AM. Based in this scenario answer the following questions.

1.	a)	Describe appropriate data preprocessing techniques for the sensor-record issues identified in the scenario, with justification for each choice to ensure data quality.	[3]	CO1
	b)	Explain the impact of the engineer's encoding approach, a more suitable alternative with justification, and the resulting encoding of all three motor types.	[3]	
	c)	Explain your choice of input features and a suitable Machine Learning model for motor failure prediction based on the scenario.	[4]	

A delivery company wants to predict whether an order will arrive on time. The company has collected the following historical delivery records.

ID	Route zone	Traffic	Distance (km)	On time
1	Central	Heavy	2	Yes
2	Central	Heavy	6	Yes
3	Outer	Light	2	Yes
4	Outer	Heavy	4	No
5	Outer	Heavy	6	No
6	Remote	Light	4	Yes
7	Remote	Light	4	No
8	Remote	Light	6	Yes

The company aims to construct a decision tree using ID3, with On time as the target attribute. Each ID represents a separate delivery and must not be used as a predictor.

The company has also received ~~two~~¹ new delivery requests with unknown outcomes:

Request	Route zone	Traffic	Distance (km)
R1	Outer	Heavy	5

Based on the dataset, answer Questions 2(a) and 2(b).

2.	a)	Apply the ID3 algorithm to the given dataset to construct a decision tree, with entropy and Information Gain calculations for root selection and subsequent splits, candidate numerical thresholds, branch conditions, and leaf-node decisions, including tie resolution.	[10]	CO2																		
	b)	Use the constructed decision tree to predict whether each new delivery will arrive on time, with the corresponding decision path.	[1]																			
	c)	<p>A manufacturing company uses a hard-margin SVM to classify products as Defective (+1) or Acceptable (−1). Each product is represented by three sensor features, $(x=(x_1,x_2,x_3))$. The sensor readings and optimized Lagrange multipliers are given below.</p> <table border="1"><thead><tr><th>Product</th><th>Sensor readings $(x=(x_1,x_2,x_3))$</th><th>Class (y)</th><th>Lagrange multiplier</th></tr></thead><tbody><tr><td>A</td><td>(0, 0, 0)</td><td>−1</td><td>0.75</td></tr><tr><td>B</td><td>(−1, −1, −1)</td><td>−1</td><td>0</td></tr><tr><td>C</td><td>(2, 2, 0)</td><td>+1</td><td>0.25</td></tr><tr><td>D</td><td>(0, 0, 2)</td><td>+1</td><td>0.50</td></tr></tbody></table> <p>A new product has sensor readings $(P=(1,1,0.5))$.</p> <p>Apply the hard-margin SVM formulation to obtain the support vectors, weight vector, bias, decision boundary, full margin width, and predicted class of P, with the necessary calculations.</p>	Product		Sensor readings $(x=(x_1,x_2,x_3))$	Class (y)	Lagrange multiplier	A	(0, 0, 0)	−1	0.75	B	(−1, −1, −1)	−1	0	C	(2, 2, 0)	+1	0.25	D	(0, 0, 2)	+1
Product	Sensor readings $(x=(x_1,x_2,x_3))$	Class (y)	Lagrange multiplier																			
A	(0, 0, 0)	−1	0.75																			
B	(−1, −1, −1)	−1	0																			
C	(2, 2, 0)	+1	0.25																			
D	(0, 0, 2)	+1	0.50																			