JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

Make-up Examination-Nov-2025

COURSE CODE (CREDITS): 25B11CI512 MAX. MARKS: 25

COURSE NAME: Intelligent techniques for predictive data analytics

COURSE INSTRUCTORS: NTS* MAX. TIME: 1 Hour 30 Minutes

Note: (a) All questions are compulsory.

(b) Use of calculator is permitted.

Q.No	Question	CO	Marks
Q1	A k-Nearest Neighbour model is to be used for predicting whether a student *passes* a course. The training dataset is shown below.		6

ID	Study Hours (x_1)	Score (%) (x_2)	Class
1	2.0	45	Fail
2	4.0	70	Pass
3	5.0	60	Pass
4	7.0	85	Pass
5	3.0	50	Fail
6	6.0	80	Pass

A new student has Study Hours = 4.5 and Score = 72.

(a) In one short paragraph, explain how k-NN classifies a new instance (what role distances and k play).

(b) Produce a table showing min–max normalized values for all training rows and for the test point (range [0,1]). (Show min and max used for each feature.

(c) Using the normalized values, compute the Euclidean distance from the test point to each training example (show all distances).

(d) For **k = 3**, list the three nearest neighbours (by ID) and give the predicted class by majority voting.

Q2	A small retail dataset is shown; the target attribute is Buys (Yes/No). Compute the Information Gain for attribute Salary (treat Salary as categorical with values High/Medium/Low). Show every step: total entropy of Buys, entropy after splitting on Salary (weighted), and the final information gain (use log base 2).		6

ID	Age	Salary	Buys
1	Young	High	No
2	Middle	Medium	Yes
3	Senior	Low	No
4	Middle	High	Yes
5	Young	Medium	No

Q3	(a) A binary classifier for spam detection gave the following results on a test		4 + 2

(26)

	dataset: True Positives (TP) = 200, False Positives (FP) = 60, False Negatives (FN) = 40, True Negatives (TN) = 700. Calculate Accuracy, Precision, Recall, and F1-score (round to 3 decimal places). (b) Explain the difference between Precision and Recall. How can both metrics be affected by choosing a different classification threshold? Give a brief example in context of spam filtering.		
Q4	(a) Explain the role of a kernel function in Support Vector Machines (SVM). How does it allow the SVM to classify data that is not linearly separable? Give one commonly used kernel and its form. (b) What is over fitting in decision trees? Describe two techniques that help reduce over fitting (e.g., pruning, limiting depth, or cross-validation) and explain how each works.		4 + 3