Graphic Era HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 20..)
University under section 2(f) of UGC Act, 1956

End Term (Even) Semester Examination May-June 2025

Roll no.....

Name of the Course and semester: MCA, 2nd Sem
Name of the Paper: Machine Learning-I
Paper Code: TMC-211
Time: 3-hour Maximum Marks: 100

Note:

(i) All the questions are compulsory.
(ii) Answer any two sub questions from a, b and c in each main question.
(iii) Total marks for each question is 20 (twenty).
(iv) Each sub-question carries 10 marks.

Q1. (2X10=20 Marks)

a. "Machine learning is an integral part of Artificial Intelligence." Comment on the above statement. Explain various applications of machine learning. (CO1)

b. Explain the role of NumPy and Pandas libraries in the Machine Learning workflow. Describe with examples how NumPy assists in numerical computations and how Pandas simplifies data analysis and manipulation tasks. (CO1)

c. Explain the bias-variance tradeoff in machine learning and what is overfitting in machine learning? How can it be addressed? (CO1)

Q2. (2X10=20 Marks)

a. Explain the term Decision Tree in Machine Learning. How it is different from Random Forest. Explain the challenges that occurs in the implementation of Decision Trees, also give the solution to overcome these challenges. (CO2)

b. The height details of the boys and girls are given in table:

Height of the Boys (x_i)	65	70	75	78
Height of the Girls (y_i)	63	67	70	73

Fit a suitable line of best fit for the above data. (CO2)

c. Discuss different evaluation metrics used to assess the performance of regression models. Explain Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared score with suitable examples. (CO2)

Q3. (2X10=20 Marks)

a. Apply Apriori Algorithm on the grocery store example with support threshold s=33.34% and confidence threshold c=60%, where H, B, K, C, and P are different items purchased by customers.

(i) Show all final frequent itemsets.
(ii) Specify the association rule that rare generated
(iii) Show final association rules sorted by the confidence

Transaction ID	Items
T1	H,B,K
T2	H,B
T3	H,C,P
T4	P,C
T5	P,K
T6	H,C,P

Graphic Era HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Even) Semester Examination May-June 2025

b. Cluster the following 8 points into three clusters.

A1(2,10), A2(2,5), A3(8,4), A4(5,8), A5(7,5), A6(6,4), A7(1,2), A8(4,9).

Initial cluster centres are A1(2,10), A4(5,8), A7(1,2). Use k-means to find the three cluster centres after 2nd iteration. Use Manhattan distance to calculate the distances. (CO3)

c. Why do we use hierarchical clustering? Explain linkage criteria. Consider the similarity matrix given below:

	P1	P2	P3	P4	P5	P6
P1	1.00	0.70	0.65	0.40	0.20	0.05
P2	0.70	1.00	0.95	0.70	0.50	0.35
P3	0.65	0.95	1.00	0.75	0.55	0.40
P4	0.40	0.70	0.75	1.00	0.80	0.65
P5	0.20	0.50	0.55	0.80	1.00	0.85
P6	0.05	0.35	0.40	0.65	0.85	1.00

Show the hierarchy of clustering created by single-link technique. (CO3)

Q4. (2X10=20 Marks)

a. Consider a feed-forward neural network with the following details:

Input layer: 2 neurons (x1, x2). Hidden layer: 3 neurons (h1, h2, h3). Output layer: 1 neuron (o1)

Given the following input and biases:

Inputs: x1=0.3, x2=0.7

Biases: Bias_h1= -0.2, Bias_h2= 0.4, Bias_h3= 0.1, Bias_o1= 0.5

Assume the following weights:

w1_h1= 0.6, w1_h2= 0.3, w1_h3= 0.8

w2_h1= -0.2, w2_h2= 0.7, w2_h3= 0.9

w1_o1= 0.3, w2_o1= -0.1, w3_o1= 0.7

Calculate the output values (o1) for the given inputs using Sigmoid activation function. (CO4)

b. Explain RELU, Sigmoid and Softmax activation functions with the help of their working suitability and equations. (CO4)

c. What do you mean by CNN? Explain all the layers of CNN architecture with its working suitability. (CO4)

Q5. (2X10=20 Marks)

a. An ML model trained on an imbalanced dataset (80% male, 20% female) showed poor performance on female users. Explain how data imbalance affects fairness and model accuracy. Describe two techniques (such as SMOTE, re-sampling, fairness constraints) to address data imbalance. (CO5)

b. Describe how random search differs from grid search for hyperparameter tuning. Why is random search often preferred when the search space is large? (CO5)

c. What are the different ways to handle missing values in a dataset? (CO5)