BENNETT UNIVERSITY
THE TIMES GROUP

Enrolment No: E22CSEU0827 Name of Student: MADHAV GUPTA

Department/ School: SCSET

MID-TERM EXAMINATION, EVEN SEMESTER MARCH 2024

COURSE CODE : CSET301 **MAX. DURATION: 1 HR**
COURSE NAME : Artificial Intelligence and Machine Learning
PROGRAM : B. Tech **TOTAL MARKS: 15**

Mapping of Questions to Course and Program Outcomes										
Q. No.	1	2	3	4	5					
CO	CO1	CO1	CO2	CO2	CO1					
PO	PO1	PO1	PO2	PO1	PO3					

GENERAL INSTRUCTIONS: -

1. Do not write anything on the question paper except **name, enrolment number** and **department/school.**
2. Carrying mobile phone, smart watch and any other non-permissible materials in the examination hall is an act of **UFM.**

COURSE INSTRUCTIONS:

a) All Questions are Compulsory.
b) If require any missing data; then choose suitably
c) Draw neat diagrams wherever necessary

SECTION A Max Marks:15

1) Consider a dataset representing the relationship between the number of study hours (independent variable) and exam scores (dependent variable) for a group of students. The data points are as follows: **(4 Marks)**

Study Hours (X)	Exam Scores (Y)
2	75
3	82
4	88
5	91
6	98

Now, using the method of least squares, calculate the coefficients (slope and intercept) for the linear regression model $Y=b0+b1 \cdot X$. Demonstrate your calculations and interpret the meaning

of the slope coefficient in the context of this dataset. (Assume $b1$ is the slope, and $b0$ is the intercept).

2) Consider a dataset representing whether students pass (1) or fail (0) in an exam based on the number of hours they studied. The dataset is as follows: **(4 Marks)**
Hours Studied (X): 2, 3, 4, 5, 6
Pass (1) or Fail (0) (Y): 0, 0, 1, 1, 1
Using Logistic Regression, calculate the coefficients (intercept and slope) for the logistic regression model $Y=1/(1+e^{-(b0+b1.X)})$. Demonstrate your calculations and interpret the meaning of the slope coefficient in the context of this dataset. (Assume $b1$ is the slope, and $b0$ is the intercept).

3) Suppose you have a binary classification model that predicts whether an email is spam (positive class) or not spam (negative class). The confusion matrix for this model is as follows: **(2 Marks)**

	Actual Class: Spam	Actual Class: Not Spam
Predicted Class: Spam	150	20
Predicted Class: Not Spam	10	200

Utilize your understanding to calculate these performance metrics: Accuracy, Precision, Recall, F1 Score. Demonstrate your calculations for each metric.

4) Describe the concepts of overfitting, underfitting, bias, variance in the context of machine learning. Investigate how overfitting and underfitting relate to bias and variance. **(2 Marks)**

5) Consider the following dataset, which represents a small collection of fruits described by their features and whether they are liked (Yes) or not liked (No). You are tasked to start constructing a decision tree to predict if a fruit will be liked, using the concept of information gain as your splitting criterion: **(3 Marks)**

Fruit ID	Color	Size	Season	Liked
1	Red	Big	Summer	Yes
2	Green	Small	Winter	No
3	Yellow	Medium	Summer	Yes
4	Green	Big	Winter	No
5	Red	Small	Summer	Yes
6	Yellow	Big	Winter	No

Note: Demonstrate your calculations for each below metric with formula.

a) Calculate the Entropy of the Entire Dataset.
b) Calculate the Entropy for Each Feature Split.
c) Calculate Information Gain and Determine the Best Splitting Feature.

-ALL THE BEST-