Candidate seat No: _____

Charotar University of Science and Technology [CHARUSAT]
Faculty of Technology & Engineering
Department of Computer Science & Engineering
Subject: CS360 Machine Learning
Unit Test - I

Semester: 6th Semester **Maximum Marks: 30**
Date: 13/02/2023 (Monday) **Time: 9:30 am to 10:30 am**

Instructions:
(i) Attempt ***all*** the questions.
(ii) Figures to the right indicate ***full*** marks.
(iii) Make suitable assumptions and draw neat figures wherever if required.
(iv)Answer the following questions:

1 Suppose there are 200 covid-19 patient details are available for testing at CHARUSAT Hospital. Out of the total population, let's consider the sample of 10 patients' age is available in the following list. Find Mean, Mode, Range and Standard deviation. **[04]**

Patient Id	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Age (In Years)	27	32	33	61	43	27	27	45	55	23

2 Calculate Z-Score normalization for following data: 8, 10, 15, 20. **[04]**

OR

2 Create a boxplot for given data: 90, 100, 100, 110,110,120,130,130,140,160,210. **[04]**

3 What is Machine Learning? Which are the types of Machine Learning? **[02]**

4 Match the following. **[03]**

1. Logistic Regression	A. Prediction of car mileage
2. Linear Regression	B. Classification of customers who are eligible for loan or not
3. Lazy learning algorithm	C. Decision tree
	D. KNN

5 Answer the following set of questions based on the given data on flower varieties. **[04]**

Sr. No.	**Sepal_length**	**Sepal_width**	**Petal_length**	**Petal_width**	**Variety**
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3	1.4	0.2	Setosa
3	7		4.7	1.4	Versicolor
4	6.4	3.2	4.5	1.5	Versicolor
5	6.3	2.9		1.8	Virginica
6	6.5	3	5.8	2.2	Virginica

i. Identify the missing values present in this dataset.
ii. Write any two ways to deal with the missing values.
iii. Which attribute can be considered as a target column and feature columns?
iv. Write two algorithm names to solve this problem.

6 The following dataset represents 5 students' results by considering two subject marks of the 6th CSE class. Find the result of sample number 6th using k-NN algorithm. Consider K=3 and Euclidean Distance. **[05]**

Sr. No	Marks(Maths)	Marks(CS)	Result
1	4	3	Fail
2	6	7	Pass
3	7	8	Pass
4	5	5	Fail
5	8	8	Pass
6	6	8	?

OR

6 Suppose we have a dataset of **weather conditions** and the corresponding target variable "**Play**". We can use this dataset to decide whether we can play or not on a particular day according to the weather conditions.
Problem: If the weather is sunny, then the Player should play or not? Solve using Naïve Bayes classifier. Show your calculations. **[05]**

Sr. No	Outlook	Play
1	Rainy	Yes
2	Sunny	Yes
3	Overcast	Yes
4	Overcast	Yes
5	Sunny	No
6	Rainy	Yes
7	Sunny	Yes
8	Overcast	Yes
9	Rainy	No
10	Sunny	No
11	Sunny	Yes
12	Rainy	No
13	Overcast	Yes
14	Overcast	Yes

7 You are a marketing analyst for Shopping Mall. You have collected following data of number of advertisements vs sales. Determine the equation of the best fit straight line using Ordinary Least Squares (OLS) Method. Also, Predict the sales when number of advertising is 13. **[05]**

Number of Ad	Sales (Units)
10	25
15	28
12	26
16	30
18	36

8 There are 10,000 samples in a dataset having 12 classes. You want to use this dataset to evaluate the machine learning algorithms like KNN, Decision Tree and Naïve bayes. Which cross-validation methods will be more beneficial to use and which will not for evaluation? Why? **[03]**

OR

8 Calculate Precision, Recall and Accuracy from the following Confusion Matrix. **[03]**

	True Positive	True Negative
Predicted Positive	60	15
Predicted Negative	10	15

****