

Enrolment No: _____ Name of Student: _____

Department/ School: _____

End Term Odd Semester Makeup Examination January – 2025

COURSE CODE: CSET211	MAX. DURATION	2 HRS
COURSE NAME: Statistical Machine Learning		
PROGRAM: BTech	TOTAL MARKS	40

Mapping of Questions to Course and Program Outcomes								
Q.No.	A1	A2	A3	A4	B1	B2	B3	B4
CO	1	1	1	2	2	3	3	3
PO	1,3	1,3,12	1,3,12	1,3,12	1,3,12	1,3,12	1,3,12	1,3,12
BTL	1,2	1,2,3	2,3,4	3,4	1,2	1,2,3	2,3,4	1,2

GENERAL INSTRUCTIONS: -

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

COURSE INSTRUCTIONS:

- a) A scientific calculator is permissible in the examination hall.
- b) All questions are compulsory

SECTION A
Max Marks:19

- A1) Explain unsupervised learning with an example. **[2 Marks]**
- A2) A retail store wants to analyze its customers' purchasing habits to create targeted marketing strategies. The store has collected data on the total amount spent by 19 customers (in \$). The store management believes that customers can be grouped into two clusters: high spenders and low spenders. Use K Means clustering to group the given data {15,15,16,19,19,20,20,21,22,28,35,40,41,42,43,44,60,61,65} into the two clusters. Assume cluster centroid as m1=16 and m2=22. **[6 Marks]**

A3) Consider the following dataset that contains 3 features (Weather, Wind, and Time) and 1 dependent variable (Go Hiking) with 2 output classes. Use Information Gain to identify the best attribute to select as the root node for building a decision tree. [6 Marks]

Weather	Wind	Time	Go Hiking
Sunny	Strong	Morning	Yes
Sunny	Weak	Morning	Yes
Rainy	Strong	Afternoon	No
Rainy	Weak	Afternoon	No

A4) Identify the most suitable machine learning task (Classification, Regression, or Clustering) for the following applications. Justify your answer for each. [1+1+1+1+1=5 Marks]

- a) Predicting the sales revenue of a retail store based on historical data.
- b) Detecting spam emails based on their content and features.
- c) Grouping online shoppers based on their browsing and purchasing behavior.
- d) Forecasting the temperature of a city for the next week.
- e) Classifying images of fruits into categories such as "Apple," "Banana," and "Orange."

SECTION B

Max Marks:21

B1) The confusion matrix for the case study: whether the person is going to buy a Mobile Phone or not, is given below [1+1+1+1=4 Marks]

	Predicted: Buy	Predicted: Not Buy
Actual: Buy	456	35
Actual: Not Buy	27	800

Find the following performance measures from the given confusion matrix

- a) Accuracy
- b) Precision
- c) Recall
- d) F1-Score

B2) a) Explain the KNN Algorithm in detail [2 Marks]

- b) A fitness app studies user activity patterns to recommend tailored fitness programs. The app has collected data on user activity, where each activity is represented by two features: Activity A1 and Activity A2. Based on these features, users are categorized into two classes: Active and Inactive.

The app wants to classify a new user activity, represented as point $P = (4, 5)$, into one of the two classes using the K-Nearest Neighbours (KNN) algorithm with $k=3$. [6 Marks]

A1	A2	Class
2	3	Active
3	4	Active
5	6	Active
7	8	Inactive
8	5	Inactive
6	7	Inactive

B3) Consider the training data in the following table, where Weather, Traffic, and TimeOfDay are independent attributes, and ReachOnTime is the dependent attribute. The Weather attribute can have values "Sunny" (S), "Rainy" (R), or "Cloudy" (C). The Traffic attribute can have values "Heavy" (H) or "Light" (L). The TimeOfDay attribute can have values "Morning" (M) or "Evening" (E). The ReachOnTime attribute has two class labels: "Yes" or "No". Using the naïve Bayes classification, predict the class label for a day where Weather = Rainy, Traffic = Heavy, and TimeOfDay = Morning.

[7 Marks]

Weather	Traffic	TimeOfDay	ReachOnTime
Sunny	Heavy	Morning	No
Rainy	Heavy	Evening	No
Cloudy	Light	Morning	Yes
Sunny	Light	Morning	Yes
Rainy	Light	Evening	Yes

B4) Explain the underfitting and overfitting in detail.

[2 Marks]

