Name: **Enrolment No:**	

UPES

End Semester Examination, May 2024

Course: Machine Learning **Semester: IV**

Program: B.Tech. CSE (AIML) **Time : 03 hrs.**

Course Code: CSAI2001 **Max. Marks: 100**

Instructions: Attempt all Questions, Usage of Scientific calculator is allowed

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	What is Sigmoid function? Explain its role in logistic regression.	**4**	**2**
Q 2	Consider a Multi-Layer Perceptron with the following architecture: Input layer: 10 neurons. Hidden layer 1: 20 neurons. Hidden layer 2: 15 neurons. Output layer: 5 neurons. Calculate the total number of parameters used in this MLP	**4**	**3**
Q 3	List the key stages and their role in the information retrieval process.	**4**	**5**
Q 4	Define centroid and dissimilarity measure in the context of cluster analysis. How are these statistical measures used in clustering algorithms?	**4**	**4**
Q 5	State the difference between bagging and boosting ensemble techniques.	**4**	**3**
SECTION B **(4Qx10M= 40 Marks)**			
Q 6	Discuss the different tasks involved in data preprocessing.	**10**	**1**
Q 7	Provide a step-by-step explanation of the K-means clustering algorithm, including how the initial centroids are chosen and how clusters are updated in each iteration.	**10**	**4**
Q 8	Consider the following data set.	**10**	**3**

	<table><tr><th>Confident</th><th>Studied</th><th>Sick</th><th>Result</th></tr><tr><td>Yes</td><td>No</td><td>No</td><td>Fail</td></tr><tr><td>Yes</td><td>No</td><td>Yes</td><td>Pass</td></tr><tr><td>No</td><td>Yes</td><td>Yes</td><td>Fail</td></tr><tr><td>No</td><td>Yes</td><td>No</td><td>Pass</td></tr><tr><td>Yes</td><td>Yes</td><td>Yes</td><td>Pass</td></tr></table> Find out whether the object with attribute **Confident = Yes, Sick = No** will Fail or Pass using Naïve Bayesian classification.		
Q 9	How do we represent documents in information retrieval systems? Discuss different document representation techniques and their advantages. **OR** Explain the following terms in context of Support Vector Machine: a) Soft and Hard Margin b) Kernel c) Hyper-plane d) Marginal Distance e) Support Vectors	**10**	**5**
	SECTION-C **(2Qx20M=40 Marks)**		
Q 10	Explain the steps of Adaptive Boosting algorithm under the following conditions: (i) Use Gini impurity while selecting stump. (ii) Show one complete iteration of the algorithm using the below mentioned dataset. <table><tr><th>Chest Pain</th><th>Blocked Arteries</th><th>Patient Weight</th><th>Heart Disease</th></tr><tr><td>Yes</td><td>Yes</td><td>205</td><td>Yes</td></tr><tr><td>No</td><td>Yes</td><td>180</td><td>Yes</td></tr><tr><td>Yes</td><td>No</td><td>210</td><td>Yes</td></tr><tr><td>Yes</td><td>Yes</td><td>167</td><td>Yes</td></tr><tr><td>No</td><td>Yes</td><td>156</td><td>No</td></tr><tr><td>No</td><td>Yes</td><td>125</td><td>No</td></tr><tr><td>Yes</td><td>No</td><td>168</td><td>No</td></tr><tr><td>Yes</td><td>Yes</td><td>172</td><td>No</td></tr></table> **OR** Use the below mentioned data set to explain the steps of Gradient Boosting algorithm for regression problem under the following conditions:	**20**	**3**

<table>
<tr><td></td><td>(i) Can restrict up to 4 leaf nodes.
(ii) Create two trees after the first single leaf tree):

<table><tr><th>Height</th><th>Favourite Color</th><th>Gender</th><th>Weight (Need to be predicted)</th></tr><tr><td>1.6</td><td>Blue</td><td>Male</td><td>88</td></tr><tr><td>1.6</td><td>Green</td><td>Female</td><td>76</td></tr><tr><td>1.5</td><td>Blue</td><td>Female</td><td>56</td></tr><tr><td>1.8</td><td>Red</td><td>Male</td><td>73</td></tr><tr><td>1.5</td><td>Green</td><td>Male</td><td>77</td></tr><tr><td>1.4</td><td>Blue</td><td>Female</td><td>57</td></tr></table></td><td></td><td></td></tr>
<tr><td>Q 11</td><td>Discuss the different types of regression models including simple linear regression, multiple linear regression, polynomial regression, and logistic regression. Provide mathematical formulations for each regression model, outlining the assumptions, advantages, and limitations of each.</td><td>20</td><td>1</td></tr>
</table>