U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: VI**
Branch: Electronics and Communication Engineering **Duration: 3 hrs.**
Course Code: 22EC6PE2ML **Max Marks: 100**
Course: Machine Learning

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

				CO	*PO*	Marks
			UNIT – I			
1	a)		Differentiate supervised, unsupervised and reinforcement ML techniques. Identify the type of machine learning to be applied for the following use cases. Justify your answers. i. Rainfall prediction ii. An intelligent agent navigates the vehicle iii. Spam detection in emails iv. Image segmentation in health care	*CO* 2	*PO2*	**10**
	b)		Create a data frame with four columns and four rows from lists using Pandas. Check and analyze the following. i. Dimension ii. Get the data types iii. Descriptive statistics iv. Select the first three rows in the data frame v. Drop a column vi. Add a column	*CO* 3	*PO3*	**10**
			OR			
2	a)		Analyze the measure of central tendency and measure of dispersion of statistical data analysis.	*CO* 2	*PO2*	**10**
	b)		Using Numpy and Pandas library functions, elaborate the utilities of linspace(), arrange(), randint(), isna(), info(), Describe() with example.	*CO* 3	*PO3*	**10**
			UNIT – II			
3	a)		Consider a scenario where we have given a medical data about some patients. The data contains the information of the blood pressure for a patient along with his/her age. Find the estimators for the best fit line using OLS method for the given data.	*CO* 1	*PO 1*	**08**

Age in Years(X)	Blood Pressure(y)
37	127
49	135
72	160
84	180
20	110

<table>
<tr><td></td><td></td><td>b)</td><td>Given a dataset, outline the steps to implement gradient descent to train a linear regression model. Analyze the potential impact of learning rate on the convergence of the gradient descent algorithm.</td><td>CO 2</td><td>PO2</td><td>08</td></tr>
<tr><td></td><td>c)</td><td>Discuss what might happen if the learning rate is too high or too low with the support of a quadratic curve.</td><td>CO 2</td><td>PO2</td><td>04</td></tr>
<tr><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td>4</td><td>a)</td><td>Consider the following dataset where the target variable is the actual house prices and the predicted house prices are given by a regression model:
<table>
<tr><td>Actual Price of a house (in Lakh)</td><td>380</td><td>180</td><td>400</td><td>650</td><td>890</td></tr>
<tr><td>Predicted price of a house (in Lakh)</td><td>410</td><td>220</td><td>380</td><td>590</td><td>910</td></tr>
</table>
Calculate the following evaluation metrics for the given dataset:
1) Mean Absolute Error (MAE), 2) Root Mean Squared Error (RMSE), 3) Relative Squared Error (RSE), 4) R^2-score.</td><td>CO 1</td><td>PO1</td><td>08</td></tr>
<tr><td></td><td>b)</td><td>Discuss the role of regularization (L1 and L2) in machine learning models. Compare and contrast their effects on model complexity and performance.</td><td>CO 2</td><td>PO2</td><td>08</td></tr>
<tr><td></td><td>c)</td><td>Explain the scenario of a model suffering from over fit and under fit. How is regularization used to overcome it?</td><td>CO 2</td><td>PO2</td><td>04</td></tr>
<tr><td></td><td></td><td>UNIT - III</td><td></td><td></td><td></td></tr>
<tr><td>5</td><td>a)</td><td>Design a classification model using Support Vector Machine (SVM). Discuss the steps involved in building SVM classifier.</td><td>CO 3</td><td>PO3</td><td>10</td></tr>
<tr><td></td><td>b)</td><td>Design a Decision Tree structure for the dataset below based on the information gain for the features “fever” and “congestion”. Indicate which is the best split root node the model can select.
<table>
<tr><td>Sore throat</td><td>Fever</td><td>Swollen glands</td><td>Congestion</td><td>Diagnosis</td></tr>
<tr><td>Yes</td><td>Yes</td><td>Yes</td><td>Yes</td><td>Covid</td></tr>
<tr><td>No</td><td>No</td><td>No</td><td>Yes</td><td>Allergy</td></tr>
<tr><td>Yes</td><td>No</td><td>Yes</td><td>No</td><td>Covid</td></tr>
<tr><td>No</td><td>Yes</td><td>No</td><td>Yes</td><td>Cold</td></tr>
<tr><td>Yes</td><td>No</td><td>No</td><td>Yes</td><td>Allergy</td></tr>
<tr><td>No</td><td>Yes</td><td>No</td><td>Yes</td><td>Cold</td></tr>
<tr><td>Yes</td><td>Yes</td><td>No</td><td>Yes</td><td>Cold</td></tr>
</table></td><td>CO 3</td><td>PO3</td><td>10</td></tr>
</table>

			OR			
	6	a)	Describe the step-by-step process of k-Nearest Neighbors (kNN) algorithm Explain the significance of distance measures such as Euclidean and Manhattan in the kNN algorithm. Analyze how the choice of distance measure can influence classification outcomes, providing examples to support your explanation. With X1 = 7 and X2 = 5. Guess the classification of new unseen data. (K=3) Feature 1 (X1) | Feature 2 (X2) | Y 9 | 6 | FALSE 8 | 4 | TRUE 3 | 2 | FALSE 6 | 3 | FALSE 12 | 8 | TRUE	***CO*** 2	***PO2***	**10**
		b)	Differentiate between the concepts and techniques of ensembling in machine learning, specifically focusing on Bagging and Boosting, and explain how Random Forest utilizes these techniques.	***CO*** 2	***PO 2***	**10**
			UNIT – IV			
	7	a)	Define the Naive Bayes algorithm. Given a small dataset, demonstrate how to classify a new instance using the Naive Bayes algorithm. Show your calculations step-by-step.	***CO*** 2	***PO2***	**08**
		b)	Differentiate between logit() & sigmoid() functions in the context of logistic regression	***CO*** 1	***PO1***	**06**
		c)	Compare and contrast the associations between agglomerative clustering and divisive Clustering	***CO*** 1	***PO1***	**06**
			OR			
	8	a)	Derive the process of optimizing the weight parameters with appropriate cost function using Logistic Regression algorithm.	***CO*** 1	***PO1***	**10**
		b)	With a sample dataset, elaborate the Naïve Bayes algorithm for a binary classifier and mention the Naïve assumptions.	***CO*** 2	***PO2***	**10**
			UNIT – V			
	9	a)	Define Deep Learning and discuss two of its applications. Illustrate a single layer perceptron model.	***CO*** 2	***PO2***	**08**

		b)	For the given dataset., apply k-means clustering(k=3) for one iteration and write the cluster assignments. Assume Euclidean distance measure. Consider the 3rd , 6th and 10th data points as the initial centroids.	***CO*** 2	***PO2***	**12**

Data point	X	Y
1	1.9	0.97
2	1.76	0.84
3	2.32	1.63
4	2.31	2.09
5	1.14	2.11
6	5.02	2.8
7	5.74	3.4
8	2.25	3.9
9	6.71	3.7
10	4.15	4.2

			OR			
	10	a)	With a neat diagram of multilayer perceptron, explain how an artificial neural network works.	***CO*** 2	***PO2***	**10**
		b)	Analyze different types of clustering used for unsupervised machine learning.	***CO*** 2	***PO2***	**10**

****