Roll No. | | | | | | | | | | | |

PRESIDENCY UNIVERSITY
BENGALURU
School of Engineering
Mid - Term Examinations - November 2024

Semester: V

Course Code: ECE3025

Course Name: Artificial Intelligence with Python

Program:B.Tech

Date: 07/11/2024

Time: 11.45am to 01.15pm

Max Marks: 50

Weightage: 25%

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks.   **5Qx2M=10M**

1	A confusion matrix is used in evaluating the performance of a classification model. Describe "precision" and "recall" with an example.	**2 Marks**	**L1**	**CO1**
2	Labeled data and unlabeled data are used in machine learning. Describe both and where they will be used to perform any task.	**2 Marks**	**L1**	**CO1**
3	Deep learning and Machine Learning are under the umbrella of artificial intelligence. How does deep learning differ from machine learning?	**2 Marks**	**L1**	**CO1**
4	Metrics are used to evaluate the performance of any system. Mention the metrics for evaluating unsupervised learning algorithms for optimal distance between clusters and data points.	**2 Marks**	**L1**	**CO2**
5	Random forest algorithm is extension of decision tree algorithm. How does a random forest model work as a classifier and regression?	**2 Marks**	**L1**	**CO2**

Part B

Answer ALL Questions. Each question carries 10 marks.   **4QX10M=40M**

6	a.	Data preprocessing plays a crucial role in machine learning. Let us assume that this is the data set we are working with ([-1.9, 2.3, 2.1, 0.8], [-2.9, 1.5,-0.8, 2.0], [3.2,-2.5, 2.7, 0.9]). This data set needs to be binarized with a threshold of 2.5 for preprocessing. Then what is binarized data for the given data set?	**5Marks**	**L2**	**CO1**

b.	Label encoding refers to the process of transforming the word labels into numerical form. How does label encoding perform if the input labels i) input_labels = ['red', 'black','red', 'green', 'black', 'yellow', 'white'] and ii) input_labels = (1, 2, 0, 3).	**5Marks**	**L2** **CO1**

Or

7 a. Assume the classifier for odds of passing course Log (odds)= -64+2* hours. the given data sets are as **5Marks** **L2** **CO1**

Hours	Pass(1)/Fail(0)
29	0
15	0
33	1
28	1
45	1

(i) Calculate the probability of pass who studied 32 hours.

(ii) How many hours student should study that makes he will pass the course with the probability of more than 90%.

b. Consider the provided confusion matrix. Calculate the specified performance metrics. **5Marks** **L2** **CO1**
(i) Accuracy
(ii) Precision
(iii) Recall
(iv) F1 Score

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

8 A Naïve Bays classifier is a machine learning algorithm that is based on Bayes' theorem. It is a probabilistic classifier that assumes independence between features. The Naïve Bayes classifier is commonly used for text classification tasks such as spam detection and sentiment analysis. Apply the Naïve Bays classifier, Consider the hypothesis: whether the person plays tennis under the observation of a sunny **10Marks** **L2** **CO1**

Use Naïve Bays classifier, Consider the hypothesis is whether the person play the tennis under the observation of sunny

Outlook	Sunny	Sunny	Overcast	Rain	Rain	Rain	Overcast	Sunny	Sunny	Rain	Sunny	Overcast	Overcast	Rain
Play Tennis	No	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No

Or

9	How do you draw the 'HYPERPLANE' in given data points? The points (4,1),(4,-1) and (6,0) are belongs to positive class and points (1,0),(0,1) and (0,-1) are belongs to negative class.	10Marks	L2	CO1

10	Consider all the features, and play tennis as the target feature. Check whether the person will play tennis on the particular day using decision tree classifier.	10Marks	L2	CO2

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Or

11	Consider a scenario where we want to predict the price of a house with 1900 sq ft, 3 bedrooms, and located 6 km from the city center. Apply random forest regression and compare the results with decision tree regression.	10Marks	L2	CO2

12	The aim of the K-means algorithm is to partition a given dataset into K clusters, where each data point belongs to the cluster with the nearest mean. This algorithm aims to minimize the within-cluster variance and maximize the between-cluster variance. Additionally, it iteratively assigns data points to clusters and updates the cluster means until convergence is achieved. Find the optimized centroid points for the cluster having 8-points (with (x,y) representing locations) in three clusters: A1 (2, 10), A2 (2, 5), and A3(8, 4), A4(5,8), A5(7,5), A6(6,4), A7(1,2), and A8(4,9). Initial cluster centers are A1 (2, 10), A4 (5, 8), and A7 (1, 2). The distance function between two points a=(x1, y1) and b=(x2, y2) is defined as: $\rho(a, b) = |x2 - x1| + |y2 - y1|$.	10Marks	L2	CO2

Or

13   Imagine you're working for an online retail company and your goal is to segment customers based on their purchasing behavior. You plan to use **K-means clustering** to group customers, but before you begin, you need to figure out the optimal number of clusters (i.e., the number of customer segments). However, determining the optimal number of clusters is a critical step. What are the appropriate techniques for selecting the right number of clusters, and how can they be explained with a practical scenario?   **10Marks   L2   CO2**