|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

 ****

**Presidency University**

**Bengaluru**

 **SCHOOL OF COMMERCE**

**Summer Term End Term Examinations, August 2024**

**Winter Semester**: 2023 - 24

**Course Code**: ECE3106

**Course Name**: Introduction to Data Analytics

**Program & Sem**: BBB, BBE & VIth Sem

**Date**: 12/ August / 2024

**Time**: 1.00 PM To 4.00 PM

**Max Marks**: 100

**Weightage**: 50%

 **Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Question paper consists of three parts.*
3. *Scientific and Non Programable Calculators are Permitted.*
4. *Do not write any information on the question paper other than roll number.*

**Part A**

**Answer any FIVE Questions. (5 Q x 2 M = 10 M)**

1. A list can store different elements of same type. We can use different commands to display the elements of list. Write the correct output of the command- list (range (10)) (C.O 1) [Knowledge]

2. There are different functions available in python for accessing a particular element in a list. Identify the correct output of the command bisect. bisect (c, 2) to the list c = [1, 2, 2, 2, 3, 4, 7]

 (C.O.1) [Knowledge]

3. List is an array of elements of same type. We can use different commands to display the elements of list. Write the correct Output of the command list (range (0, 24, 2))

 (CO 2) [Knowledge]

4. A continuous Random Variable, X, follows a Uniform Distribution so that the probability of any value between 2 and 5 is p. What is the value of p? (C.O..3) [Knowledge]

5. List out different analytical and statistical techniques data scientists commonly use?

 (CO-3)[Knowledge]

6. In machine learning, what is the primary difference between supervised and unsupervised learning. (C.O.4) [Knowledge]

7. In KNN classification algorithm, What does K stand for ? (C.O.4) [Knowledge]

**Part B**

**Answer any FIVE Questions. (5 Q x 10 M = 50 M)**

8. You can select sections of most sequence types by using slice notation, which in its basic form consists of start: stop passed to the indexing operator []. For the given sequence seq = [2, 10, 7, 2, 3, 7, 5, 6, 0, 1], find out the following sequences:

a) seq[1:5] and seq[3:4]

b) seq[:8] and seq[4:]

c) seq[-3:] and seq[-7:-2]

d) seq[::3] and seq[::-1] (CO-1)[Comprehension]

9. In linear algebra, the transpose of a matrix is an operator which flips a matrix over its diagonal; that is, it switches the row and column indices of the matrix A by producing another matrix, often denoted by Aᵀ. (i) Find the transpose of the matrix A = np. arrange (2, 26, 2).reshape((3, 4))

(ii) Find the result of x=np.dot (A, A.T)  (5+5)(CO-2) [Comprehension]

10. For the data given below, show how you can use box plot to plot the data representing Quartile 1, Quartile 2, and Quartile 3.

 5, 7, 4, 4, 6, 2, 8 (CO-2) [Comprehension]

11. A random variable is a mathematical formalization of a quantity or object which depends on random events. It is a mapping or a function from possible outcomes in a sample space to a measurable space. Four coins are tossed. If Y represents the number of tails, what is P(Y ≤ 1)? List all the outcomes. Write the Sample Space. What is the probability of Y=1, Y=2 and Y=3

 (CO-3)[Comprehension]

12. A machine produces electrical components. Centered on the mean, 99.7% of the components have lengths between 1.176 cm and 1.224 cm. Assuming this data is normally distributed, what are the mean and standard deviation? (CO-3)[Comprehension]

13. Classification algorithms can be divided into Linear and Non linear models. The goal of the an algorithm is to create the best line or decision boundary (Hyperplane) that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. Identify the algorithm and Illustrate the same with neat diagram.

 (CO-4) [Comprehension]

14. The Data set of Pass or Fail of 5 students in an exam is given in table below.

|  |  |
| --- | --- |
|  Hours of study | Pass (1)/Fail(0) |
| 29 | 0 |
| 33 | 1 |
| 15 | 0 |
| 28 | 1 |
| 39 | 1 |

Use Logistic Regression as the classifier to calculate the probability of Pass of the student who studied 28 hours. Assume the model suggested by the optimizer for odds of passing the course is *log (odds) = -64+(2 x hours)*. (C.O.4) [Comprehension]

**Part C**

**Answer any TWO Questions. (2 Q x 20 M = 40 M)**

15. a) Data scientists spend close to 75% of their time in analyzing data and engineering features which are indeed a difficult and time-consuming processes. From the given Table Identify the following – Categorical data, Numerical data, ordinal data, Discrete and continuous data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Section | Marks  | Rank  | Percentage  | grade |
| Anusha | 1 | 34 | 4 | 74 | C |
| Anupama | 2 | 48 | 1 | 96 | A |
| Vijay | 3 | 47 | 2 | 94 | A |
| Pritam | 2 | 38 | 3 | 76 | B |

 (10)(CO-1)[Application]

b) File handling is an important part of any web application. Python has several functions for creating, reading, updating, and deleting files .

You have a file named “demo.txt” located in “C:/Users/lovely/desktop” in your computer. Write python instructions to perform the following operations

i. Open the file and read all the characters in the file

ii. Open the file and read only first 10 the characters in the file

iii. Open the file and read all the lines in the file

iv. Open the file and read a line in the file

v. Without deleting the data in the file add your name and your ID number.

 (10)(CO-2)[Application]

16. Mean and variance is a measure of central dispersion. Mean is the average of given set of numbers. The average of the squared difference from the mean is the variance. Ten friends scored the following marks in their end-of-year math exam:

23%, 37%, 45%, 49%, 56%, 63%, 63%, 70%, 72% and 82%

Find the mean, median, mode, variance (population), standard deviation. (CO-3)[Application]

17. Supervised Machine learning algorithm always produces Categorical output for the labled input data. Using K- Nearest Neghbours Classification algorithm, find the class of the new data in the last row of the table given below.



 (C.O.4) [Applcation]