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



**PRESIDENCY UNIVERSITY**

**Bengaluru**

|  |
| --- |
| **End - Term Examinations – JANUARY 2025** |
| **Date:** 04 - 01- 2025 **Time:** 9:30 am – 12:30 pm |

|  |  |  |
| --- | --- | --- |
| **School:** School of Engineering | **Program:** B.Tech ECE | |
| **Course Code :** ECE3106 | **Course Name :** Introduction to Data Analytics | |
| **Semester**: VII | **Max Marks**: 100 | **Weightage**: 50% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CO - Levels** | **CO1** | **CO2** | **CO3** | **CO4** |
| **Marks** | **14** | **14** | **36** | **36** |

**Instructions:**

1. *Read all questions carefully and answer accordingly.*
2. *Do not write anything on the question paper other than roll number.*

**Part A**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Answer ALL the Questions. Each question carries 2marks. 10Q x 2M=20M** | | | | |
| **1** | Summarize the analytical and statistical techniques data scientists commonly use? | **2 Marks** | **L1** | **CO1** |
| **2** | Python has become one of the most popular interpreted programming languages, along with Perl, Ruby, and others. What is the name assigned for these programming languages. | **2 Marks** | **L1** | **CO1** |
| **3** | There are different ways to sort the elements of a list. Identify the correct input command for sorting the list  b = ['saw', 'small', 'He', 'foxes', 'six’] by their lengths. | **2 Marks** | **L2** | **CO2** |
| **4** | While defining a two-dimensional array, we need to define the dimension, size and shape of a matrix.  For the matrix b= np. array ([[1, 2], [3, 4]]), find its dimension, size and shape | **2 Marks** | **L1** | **CO2** |
| **5** | 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? | **2 Marks** | **L1** | **CO3** |
| **6** | Compute the Median of the sequence of numbers 2, 4,7, 9, 10, 14 | **2 Marks** | **L2** | **CO3** |
| **7** | Sketch the Box plot if the Quartiles in order are Q1, Q2 and Q3. | **2 Marks** | **L2** | **CO3** |
| **8** | Describe the data type at the output of classification in Supervised learning. | **2 Marks** | **L1** | **CO4** |
| **9** | Generalize the Confusion matrix in its relevent form. | **2 Marks** | **L1** | **CO4** |
| **10** | Outline the significance of K in KNN classification algorithm. | **2 Marks** | **L1** | **CO4** |

**Part B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Answer the Questions Total 80 Marks.** | | | | | |
| **11.** | **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. Give the justification for your answers.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Mobile Type | Ward No. | No. of users | Rank | Performance Percentage | Customer feedback | | Samsung | 1 | 84 | 4 | 74 | Good | | Apple | 2 | 148 | 1 | 96 | Outstanding | | LG | 3 | 147 | 2 | 94 | Excellent | | One Plus | 4 | 138 | 3 | 76 | Very good | | **10 Marks** | **L2** | **CO1** |
|  | **b.** | 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 | **10 Marks** | **L2** | **CO2** |
| **or** | | | | | |
| **12.** | **a.** | Once the problem has been defined, the model will be developed. Before deployment, the data validation phase plays an important role in the process of data analytics. Consider any real life situation as example and explain data validation process. | **10 Marks** | **L3** | **CO1** |
|  | **b.** | 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], determine 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] | **10 Marks** | **L2** | **CO2** |
|  |  |  |  |  |  |
| **13.** | **a.** | 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%  Determine the mean, median, mode, variance (population), standard deviation. | **20 Marks** | **L3** | **CO3** |
| **or** | | | | | |
| **14.** | **a.** | In probability, a real-valued function, defined over the sample space of a random experiment, is called a random variable. Two dice are tossed. The Random Variable is The sum of the scores on the two dice.  i) What is the sample space?  ii)What is the probability that the sum of the scores is 5, 6, 7 or 8 or In other words What is  P(5 ≤ X ≤ 8)?  iii) If P(X=x) = 1/12, what is the value of x? | **20 Marks** | **L3** | **CO3** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **15.** | **a** | 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 | **10 Marks** | **L3** | **CO3** |
|  | **b** | If you receive an average of two assignments per month from your professor, what is the probability that you will receive exactly one assignment from your professor in March? Assume that the number of assignments per day follows a Poisson distribution. Mention the three important constraints in Poisson distribution. | **10 Marks** | **L2** | **CO4** |
| **Or** | | | | | |
| **16.** | **a.** | The mean December evening temperature in Bangalore is 16°C and the standard deviation is 3°C. Assuming this data is normally distributed, how many days in December would you expect the evening temperature to be between 19°C and 22°C? | **10 Marks** | **L3** | **CO3** |
|  | **b.** | Consider 100 mobile phones out of which 20 are faulty. A model has been developed to check all the mobile phones. The model has predicted 62 mobile phones as working and 13 as not working. Develop the Confusion matrix for this data and write the Python code for the same. | **10 Marks** | **L4** | **CO4** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **17.** | **a.** | Design a classfier using SVM algorithm for the data points given below  (4,1), (4,-1) and (6,0) - Positive class  (1,0), (0,1) and (0, -1) - Negative class  Write Python code for the same using scikit learn. | **20 Marks** | **L4** | **CO4** |
| **Or** | | | | | |
| **18.** | **a.** | Consider weather conditions to play the sport.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Day** | **Outlook** | **Temp** | **Humidity** | **Windy** | **Play** | | 1 | Sunny | Hot | High | False | No | | 2 | Overcast | Hot | High | True | Yes | | 3 | Rainy | Mild | High | False | Yes |   Using Naïve Bayes algorithm, predict the playing condition if it is rainy, mild temp, high humidity and not windy. Write the Python code for the same. | **20 Marks** | **L4** | **CO4** |

**\*\*\*\*\* BEST WISHES \*\*\*\*\***