Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – MAY/JUNE 2025

Date: 06-06-2025 **Time:** 01:00 pm – 04:00 pm

School: SOE	**Program:** EEE / CIVIL	
Course Code: ECE3106	**Course Name:** Introduction to Data Analytics	
Semester: IV	**Max Marks**: 100	**Weightage**: 50%

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

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. **10Q x 2M=20M**

1	**Find the syntax error in the following function** **print('you're here to learn Python')**	**2 Marks**	**L1**	**CO1**
2	**Define data analytics.**	**2 Marks**	**L2**	**CO1**
3	**Compare Interpreter and Compiler.**	**2 Marks**	**L2**	**CO2**
4	**List out the various Python library functions.**	**2 Marks**	**L1**	**CO2**
5	**Mention the Sample space when the die is thrown once.**	**2 Marks**	**L1**	**CO3**
6	**Compare Uniform distribution with Normal distribution.**	**2 Marks**	**L1**	**CO3**
7	**The outlier in the following sequence is _____** **1, 2, 3, 4,5, 6, 7, 10, 14, 20, 104**	**2 Marks**	**L2**	**CO3**
8	**Compare supervised with unsupervised learning with an example.**	**2 Marks**	**L1**	**CO4**
9	**In the KNN classification algorithm, what does K stand for?**	**2 Marks**	**L1**	**CO4**
10	**Illustrate the 2D Hyperplane in the SVM classifier with its equation.**	**2 Marks**	**L1**	**CO4**

Part B

Answer the Questions **Total 80 Marks.**

<table>
<tr><td>11.</td><td>a.</td><td>Data is a collection of facts, information, and statistics, and this can be in various forms such as numbers, text, sound, images, or any other format. Explain the different types of data with an example.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>Data scientists spend close to 75% of their time analyzing data and engineering features, which is indeed a difficult and time-consuming process. From the given Table, identify the data types.
<table>
<tr><th>Mobile Type</th><th>Ward No.</th><th>No. of users</th><th>Rank</th><th>Performance Percentage</th><th>Customer feedback</th></tr>
<tr><td>Samsung</td><td>1</td><td>84</td><td>4</td><td>74</td><td>Good</td></tr>
<tr><td>Apple</td><td>2</td><td>148</td><td>1</td><td>96</td><td>Outstanding</td></tr>
<tr><td>LG</td><td>3</td><td>147</td><td>2</td><td>94</td><td>Excellent</td></tr>
<tr><td>One Plus</td><td>4</td><td>138</td><td>3</td><td>76</td><td>Very good</td></tr>
</table></td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
</table>

or

<table>
<tr><td>12.</td><td>a.</td><td>Scope in Python refers to the region of the code where a variable is accessible. Explain the LEGB rule in the scope of Python with an example.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>Data scientists spend close to 75% of their time analyzing data and engineering features which are indeed a difficult and time-consuming process. From the given Table. Identify the data types.
<table>
<tr><th>Fever Type</th><th>Ward No.</th><th>No. of patients</th><th>Rank</th><th>Percentage of people affected</th><th>Alert range</th></tr>
<tr><td>Malaria</td><td>1</td><td>84</td><td>4</td><td>74</td><td>ORANGE</td></tr>
<tr><td>Typhoid</td><td>2</td><td>148</td><td>1</td><td>96</td><td>RED</td></tr>
<tr><td>Viral Fever</td><td>3</td><td>147</td><td>2</td><td>94</td><td>RED</td></tr>
<tr><td>Dengue</td><td>4</td><td>138</td><td>3</td><td>76</td><td>YELLOW</td></tr>
</table></td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
</table>

13.	**a.**	**A probability is a number that reflects the chance or likelihood that a particular event will occur. Probabilities can be expressed as proportions that range from 0 to 1, and they can also be expressed as percentages ranging from 0%**	**20 Marks**	**L3**	**CO3**

		to 100%. **For the data given below, determine the following:** **5, 7, 4, 4, 6, 2, 8** **i)Mean, Median, Mode, Range, Quartile 1, Quartile 2, Quartile3, Interquartile range** **ii) Draw the Box plot and represent the range, Range, Quartile 1, Quartile 2, Quartile3.**			

or

14.	**a.**	**A line of best fit is a straight line that minimizes the distance between it and some data. It is an output of regression analysis and can be used as a prediction tool for indicators and price movements.** **Sam found how many hours of sunshine vs how many ice creams were sold at the shop from Monday to Friday:** **Obtain an equation for the line of best fit and predict y if x=6. Write the Python code for the same.**	**20 Marks**	**L4**	**CO3**

"x" Hours of Sunshine	"y" Ice Creams Sold
2	4
3	5
5	7
7	10
9	15

15.	**a**	**If you receive an average of two phone calls per week from your relative, what is the probability that you will receive exactly one phone call from your relative on Monday? Assume that the number of phone calls per day follows a Poisson distribution. Mention the three important constraints in the Poisson distribution.**	**10 Marks**	**L3**	**CO3**
	b	**Consider the weather conditions to play the sport.** **Using the Naïve Bayes algorithm, predict the playing condition if it is rainy, mild temperature, high humidity, and not windy.**	**10 Marks**	**L4**	**CO4**

Day	Outlook	Temp	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Overcast	Hot	High	True	Yes
3	Rainy	Mild	High	False	Yes

Or

16.	**a.**	**Mean and variance measure central dispersion. The mean is the average of a 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%. What was the standard deviation of their marks?**	**10 Marks**	**L3**	**CO3**
	b.	**Find the equation for the hyperplane (SVM) for the data points (4, 1), (4,-1), (6,0) for the positive class and (1, 0), (0, 1), (0,-1) for the negative class.**	**10 Marks**	**L4**	**CO4**

17.	**a.**	**Find the linear regression equation for the following data.** **x** 2 4 6 8 **y** 3 7 5 10 **y is the dependent variable, and x is the independent variable. Find the Mean squared error and Root mean squared error.**	**20 Marks**	**L4**	**CO4**

Or

18.	**a.**	**The table represents the data set with the details of attendance and marks and corresponding student's result in the exam.** **Attendance %** / **Marks** / **Result** 60 / 55 / Pass 76 / 67 / Pass 50 / 23 / Fail 45 / 18 / Fail 97 / 79 / Pass 73 / 24 / Fail **88 (New data)** / **87 (New data)** / *** **Apply the KNN algorithm to find the class for the new data in the above table.**	**20 Marks**	**L4**	**CO4**