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



 **PRESIDENCY UNIVERSITY**

  **Bengaluru**

|  |
| --- |
| **End - Term Examinations – JANUARY 2025** |
| **Date:** 13-01-2025 **Time:** 09:30 am – 12:30 pm |

|  |  |
| --- | --- |
| **School:** SOCSE | **Program:** B. TECH (CSD) |
| **Course Code :**CSE2028 | **Course Name:** STATISTICAL FOUNDATION FORDATA SCIENCE |
| **Semester**: V | **Max Marks**:100  | **Weightage**: 50% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CO - Levels** | **CO1** | **CO2** | **CO3** | **CO4** | **CO5** |
| **Marks** |  **24** | **24** | **24** | **28** | **NA** |

**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** | Define population and sample in statistics | **2 Marks** | **L1** | **CO1** |
| **2** | What is the difference between quantitative and categorical variables?  | **2 Marks** | **L2** | **CO1** |
| **3** | Define linear regression. | **2 Marks** | **L1** | **CO2** |
| **4** | What is spline regression, and why is it used? | **2 Marks** | **L2** | **CO2** |
| **5** | Define *classification trees*. | **2 Marks** | **L1** | **CO3** |
| **6** | Differentiate between hierarchical and K-means clustering. | **2 Marks** | **L2** | **CO3** |
| **7** | Define deep learning. | **2 Marks** | **L1** | **CO4** |
| **8** | What is a feed-forward neural network? | **2 Marks** | **L2** | **CO4** |
| **9** | Mention two applications of CNNs | **2 Marks** | **L1** | **CO4** |
| **10** | List two real-world applications of sentiment analysis. | **2 Marks** | **L3** | **CO4** |

**Part B**

|  |
| --- |
| **Answer the Questions Total 80 Marks.** |
| **11.** | **a.** | i). Analyze the similarities and differences between sample surveys, experiments, and observational studies. Discuss their real-world applications with examples, focusing on their respective strengths and weaknesses. (15M)ii) Explain why observational studies are considered less reliable than experiments.(5M) | **20 Marks** | **L4** | **CO1** |
| **or** |
|  **12.** | **a.** | i).Explain the role and importance of descriptive statistics in data analysis. Discuss how measures such as central tendency, variability, and graphical representation help summarize and interpret data in real-world contexts. (15M)i) Illustrate the significance of graphical representation in descriptive statistics. (5M) | **20 Marks** | **L4** | **CO1** |
|  |  |  |  |  |  |
| **13.** | **a.** | i)Explain the steps involved in performing linear regression analysis. Use a relevant dataset to demonstrate data preparation, model building, coefficient calculation, and result interpretation.(15M)ii). Explain how to interpret the regression coefficients in linear regression. (5M) | **20 Marks** | **L3** | **CO2** |
| **or** |
|  **14.** | **a.** | i).Compare and contrast multiple linear regression and polynomial regression. Discuss the equations, key differences, and appropriate use cases for each type of regression model. (15M)ii).Explain when it is appropriate to use polynomial regression over multiple linear regression. (5M) | **20 Marks** | **L4** | **CO2** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **15.** | **a.** | i).Explain the k-Nearest Neighbors (KNN) algorithm with an example. Analyze its advantages, limitations, and situations where it performs well. (15M)ii). Discuss when the k-Nearest Neighbors (KNN) algorithm is most effective and when it may not be suitable.(5M) | **20 Marks** | **L4** | **CO3** |
| **Or** |
| **16.** | **a.** | i).Discuss Naive Bayes classifiers and their role in text classification, using real-world examples to highlight their applications. (15M)ii). Explain how Naive Bayes classifiers are applied in text classification with an example. .(5M) | **20 Marks** | **L3** | **CO3** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **17.** | **a.** | Develop a Python program to implement a simple linear regression model using a dataset (e.g., housing prices). Explain each step, from loading the dataset to evaluating the model's performance with appropriate metrics. | **20 Marks** | **L3** | **CO4** |
| **Or** |
|  **18.** | **a.** | Develop a Python program to implement and evaluate a Naive Bayes classifier using a dataset (e.g., spam detection). Explain the steps involved in data preprocessing, model training, and performance evaluation using metrics like accuracy and confusion matrix. | **20 Marks** | **L4** | **CO4** |

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