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



 **PRESIDENCY UNIVERSITY**

  **Bengaluru**

|  |
| --- |
| **End - Term Examinations – JANUARY 2025** |
| **Date:** 10 – 01- 2025 **Time:** 09:30 am – 12:30 pm |

|  |  |
| --- | --- |
| **School:** SOCSE | **Program:** B.TECH (CEI) |
| **Course Code :** CSE3038 | **Course Name :** Applied Data Science |
| **Semester**:V | **Max Marks**:100 | **Weightage**:50% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CO - Levels** | **CO1** | **CO2** | **CO3** | **CO4** |
| **Marks** | **26** | **26** | **24** | **24** |

**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** | List two job roles associated with data science and their primary responsibilities. | **2 Marks** | **L1** | **CO1** |
| **2** | List two considerations for ensuring data quality during analysis? | **2 Marks** | **L1** | **CO1** |
| **3** | List two differences between data and information. | **2 Marks** | **L1** | **CO1** |
| **4** | Define the purpose of data profiling in the data quality assessment process? | **2 Marks** | **L1** | **CO2** |
| **5** | State differences between forward fill and backward fill techniques for handling missing values. | **2 Marks** | **L1** | **CO2** |
| **6** | Define feature aggregation in data preprocessing. | **2 Marks** | **L1** | **CO2** |
| **7** | State two differences between lazy learners and eager learners. | **2 Marks** | **L1** | **CO3** |
| **8** | Define imbalanced classification. | **2 Marks** | **L1** | **CO3** |
| **9** | Define clustering in brief. | **2 Marks** | **L1** | **CO4** |
| **10** | State for which type of data unsupervised machine learning algorithms are applied. | **2 Marks** | **L1** | **CO4** |

**Part B**

|  |
| --- |
| **Answer the Questions Total 80 Marks.** |
| **11.** | **a)****b)** | Explain atleast five applications of data science in detail.Explain indexing in a 2-D numpy array(using python). | **20 Marks** | **L2** | **CO1** |
| **or** |
| **12.** | **a)****b)** | Explain different stages of a data science project lifecycle with diagram.Describe type, shape, dimension, size and element type of a 2-D numpy array (using python). | **20 Marks** | **L2** | **CO1** |
|  |  |  |  |  |  |
| **13.** |  | Explain different feature encoding techniques available in Python. Provide code examples for label encoding, one-hot encoding, and custom encoding, explaining their use cases. | **20 Marks** | **L2** | **CO2** |
| **or** |
| **14.** |  | Explain the concept of descriptive statistics. Discuss measures of center (mean, median, mode), measures of spread (variance, standard deviation, interquartile range), and their significance in data analysis, using Python examples. | **20 Marks** | **L2** | **CO2** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **15.** |  | Explain the decision tree algorithm, its components (root, decision nodes, leaf nodes), and key terminologies (entropy, Gini impurity, pruning). Discuss its advantages and limitations**.** | **20 Marks** | **L2** | **CO3** |
| **Or** |
| **16.** |  | Describe the working of Support Vector Machines (SVMs). Differentiate between linear and non-linear SVMs and explain the kernel trick. | **20 Marks** | **L2** | **CO3** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **17.** |  | Demonstrate visualization through following plots by assuming suitable dataset (using python):a) scatter plotb) pair plotc) count plotd) heatmap | **20 Marks** | **L3** | **CO4** |
| **Or** |
| **18.** |  | Demonstrate visualization through following plots by assuming suitable dataset (using python):a) word cloudb)scatter geo plot | **20 Marks** | **L2** | **CO4** |

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