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

****

**Presidency University**

**Bengaluru**

**Ph. D Course Work End Term Examinations – JAN-FEB 2025**

**Date**: 04-02-2025

**Time**: 9.30 AM TO 12.30 PM

**Max Marks**: 100

**Weightage**: 50%

**Semester**:

**Course Code**: CSE898

**Course Name**: Machine Learning Algorithm

**Department:** SOCSE

**Instructions:**

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

**PART A**

**Answer all the Questions. Each question carries 10 marks. (6Qx 10M= 60M)**

|  |  |
| --- | --- |
| **1.** | Discuss the importance of applying scaling techniques (e.g., scalar and robust scaling) and normalization. Provide an example of how these transformations might impact model performance. **(CO:01 BL:02)** |
| **2.** | Explain the concept of a data-generating process in machine learning. Describe how understanding the underlying process helps in creating good datasets **(CO:01 BL:02)** |
| **3.** | Explain how the k-Nearest Neighbors (KNN) algorithm works for classification. **(CO:02 BL:02)** |
| **4.** | Explain the key differences between Fuzzy C-means and Spectral Clustering algorithms. .  **(CO:02 BL:02)** |
| **5.** | Compare and contrast Self-Training and Co-Training approaches in semi-supervised learning. Provide a real-world example to support your explanation**. (CO:03 BL:02)** |
| **6.** | Explain the concept of graph-based semi-supervised learning. **(CO:04 BL:02)** |

**PART B**

**Answer all the Questions. Each question carries 20 marks. (2Qx 20M= 40M)**

|  |  |
| --- | --- |
| **7.** | Describe the characteristics of a good dataset for machine learning and explain the steps involved in preprocessing data, including scaling, normalization, and whitening. Analyze the impact of using improper scaling techniques on models such as Logistic Regression and k-Nearest Neighbours (KNN). **(CO:01 BL:02)** |
| **8.** | Describe the concept of clustering and its fundamental principles. Design a scenario where DBSCAN outperforms K-means, focusing on clusters with varying density and noise. Illustrate the advantages of DBSCAN using an example dataset**. (CO:02 BL:02)** |