Roll No												

PRESIDENCY UNIVERSITY
BENGALURU

Department of Research & Development

Mid - Term Examinations - AUGUST 2024

Odd Semester: Ph.D. Course Work
Course Code: CSE 899
Course Name: Data Analytics and Visualization
Department: SOCSE

Date: 13/08/2024
Time: 09:30am – 11:00 am
Max Marks: 50
Weightage: 25%

Instructions:

(i) Read the all questions carefully and answer accordingly.
(ii) Do not write any matter on the question paper other than roll number.

PART A (THOUGHT PROVOKING)

Answer all the Questions. Each question carries 5 marks. (4Qx 5M= 20M)

1. What are the three main types of analytics, and how do they differ? (CO1)[Understand]
2. With examples for each, differentiate between the three types of data. Also mention how they are usually stored and analyzed. (CO1)[Understand]
3. Explain the concept of cluster analysis in unsupervised learning. Compare K-means clustering with hierarchical clustering in terms of their algorithms, advantages, and limitations. Provide examples of scenarios where each method would be appropriate. (CO2)[Apply]
4. Discuss the role of regression analysis in statistical methods for data analytics. Explain the difference between linear and logistic regression. (CO2)[Understand]

PART B (PROBLEM SOLVING)

Answer all the Questions. Each question carries 10 marks. (3Qx 10M= 30M)

5. You are working on a customer segmentation project for a retail company. The dataset includes numerous features such as customer demographics, purchasing history, online behavior, and social media interactions. With so many features, your model is suffering from the curse of dimensionality, making it difficult to identify distinct customer segments.
   How would you apply dimensionality reduction techniques to address this issue? Which method (e.g., PCA, t-SNE, LDA) would you choose and why? (CO1) [Apply]
6. A healthcare provider has patient data and wants to identify patterns in the types of treatments patients receive based on their medical history, age, and other factors. Describe the steps you would take to perform a cluster analysis on this data. How would you handle any potential issues with high-dimensionality, and what would you do if the data contains noise or outliers? (CO2) [Apply]
7. A telecom company wants to analyze customer call records to discover patterns in call behavior and identify potential churners. What data mining techniques would you apply to uncover meaningful patterns in the call data? How would you preprocess the data before applying these techniques, and how would you interpret the results? (CO2) [Apply]