VIT
Vellore Institute of Technology

Final Assessment Test – November/December 2023

Course: MDI3002 - Foundations of Data Science

Class NBR(s): 4666/4667 Slot: B1+TB1

Time: Three Hours Max. Marks: 100

KEEPING MOBILE PHONE/SMART WATCH, EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE

Answer ALL Questions

(10 X 10 = 100 Marks)

1. List and explain the three core aspects of effective data analysis. Consider a real-time situation, you are joined in a gym to maintain the fitness. What are the three components for the above real-time optimization problem. Explain in detail about the various types of the three components of optimization.

2. Explain the structured thinking concept. Consider a real-time situation, when you are planning to do one project in machine learning for a capstone project. Explain in detail about the various problems you faced and what are the various steps to overcome the problems using structured problem-solving technique.

3. Consider the two dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8). Compute the principal component using PCA Algorithm.

4. a) List the various measures of spread. Explain in detail about **[6]**
   i) Standard deviation
   ii) Interquartile range
   iii) Variance

   b) Define Oversampling and Undersampling with examples and neat diagrams. **[4]** Describe in detail about the various advantages and disadvantages of Oversampling and undersampling techniques.

5. Find the inverse, rank and trace of the matrix $A=\begin{pmatrix} 4 & -2 & 1 \\ 5 & 0 & 3 \\ -1 & 2 & 6 \end{pmatrix}$ and verify any two additive properties of the matrix with one example.

6. Find the critical points of the function $f(x,y)=x^2+2y^2+2xy-4y+15$ and Identify the maximized or minimized points. Check the presence of any saddle point. Explain the three conditions of the maxima and minima identification with formulas.

7. Explain in detail about Necessary and Sufficient Conditions for an Unconstrained Optimum. Describe in detail about Particle Swarm Optimization (PSO) with detailed steps.

8. a) Consider the following commands in python and provide answers for the following: **[3]**
   i) h= "hello world" , h[3:5]= _____ Ans: ____
   ii) A= "(lmn)+", A=_____
   iii) B= re.findall("\D+","12 men,11 women") , B= _____

   b) Write a python program to print all the odd numbers less than 20 with continue and using loop (if/for/while). **[7]**

9. a) List and describe the different types of charts. Explain in detail about subcategories of categorical and hierarchical types of charts with examples. [7]

b) Consider a real-time situation, you are provided with the dataset consist of various types of noises. Explain the various types of tasks followed to handle and rectify the noise in dataset. [3]

10. a) Explain in detail about the following types of data items used in R language with specific example. [6]

i) Vector ii) Matrix iii) Data Frame iv) List

b) Explain in detail about the following elements used in weka. [4]

i) Few clusterers based packages in weka.

ii) Few instance classes and filters for numerical data in weka.

⇔⇔⇔