Roll No. ☐☐☐☐☐☐☐☐☐☐☐☐   Total No. of Pages : 02

Total No. of Questions : 07

B.Sc. (AI and ML) (Sem.–4)
DATA SCIENCE
Subject Code : UGCA-1978
M.Code : 91707
Date of Examination : 26-05-2023

Time : 3 Hrs.   Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :
 1. **SECTION-A is COMPULSORY consisting of TEN questions carrying TWO marks each.**
 2. **SECTION-B contains SIX questions carrying TEN marks each and students have to attempt any FOUR questions.**

SECTION-A

 1.   **Write briefly :**

   a) What is Data Science?

   b) What are the various sources of data?

   c) Define cross validation.

   d) What are the characterstics of visual variables?

   e) Define data formats.

   f) What is parsing of data?

   g) Differentiate between residual plot and distribution plot.

   h) What is homogeneous and heterogeneous data?

   i) Write some applications of Data Science.

   j) How can generalization error be calculated?

SECTION- B

2. What is a Data Science Life Cycle? Describe with the help of suitable diagram.

3. Why is data cleaning required? Discuss some methods of data cleaning by taking suitable examples.

4. Discuss some major data security issues.

5. What makes data visualization good? Discuss the significance of choosing appropriate visual encodings.

6. What is regression? What are the various regression models? Also give their applications.

7. Differentiate between overfitting and underfitting. Give its significance in model selection.

NOTE : Disclosure of Identity by writing Mobile No. or Marking of passing request on any paper of Answer Sheet will lead to UMC against the Student.