GITAM (Deemed to be University)
[CSEN2161]
GST/GSS/GSB/GSHS. Degree Examination

IV Semester

INTRODUCTION TO DATA SCIENCE
(Effective from the admitted batch 2021–22)

Time: 2 Hours   **Max.Marks: 30**

Instructions: All parts of the unit must be answered in one place only.

Section-A
1. Answer all Questions: **(5×1=5)**

a) Identify four primary responsibilities of data scientists in defining goals.

b) In what manner can you subdivide quantitative exploratory data analysis (EDA) into additional categories.

c) Give the formula of skewness in terms of quartiles.

d) Present two main differences between Numpy and SciPy.

e) List out 5 features of ethics in data science.

Section-B
Answer the following: **(5×5=25)**

UNIT-I

2. Examine four distinct types or formats of data commonly encountered in data science in detail with suitable examples.

OR

3. Describe various methods by which data is produced within the field of data science.

UNIT-II

4. Elaborate on the procedural stages of exploratory data analysis (EDA).

OR

5. Examine the array of Python and R libraries designed to facilitate exploratory data analysis (EDA).

UNIT-III

6. Describe the multivariate analysis of variance (MANOVA) method and provide an illustrative example to aid in understanding.

OR

7. The following data represents the scores obtained by students in a mathematics test: 60,65,70,75,80,85,90,95,100

   Calculate the Karl Pearson coefficient of skewness for the given data.

UNIT-IV

8. Write short notes on the following of the Weka?

   a) Clustering

   b) Association

   c) Attribute Selection

OR

9. Compare and highlight the differences in functionality and features between scikit-learn and Pylearn2.

UNIT-V

10. Discuss the following Data Ownership Policies.

    a) Between academic institution and industry

    b) Between authors and journals

OR

11. Explain the 5 different parties that actively take part in data ownership.