[Apr-24]

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

IV Semester

INTRODUCTION TO DATA SCIENCE

(Effective for 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) List the different classifications of flat files that are frequently encountered within the field of data science.

b) Present any two resource tools utilized for conducting data analysis.

c) Which central tendency measures can be possible for nominal and ordinal data?

d) What are the main four statistical features of R?

e) Enumerate the key components that constitute the five Cs framework within the field of data science.

Section-B

Answer the following: **(5×5=25)**

UNIT-I

2. Provide brief descriptions of the following philosophies guiding the practice of data science.

   Software Developer Vs Data Scientist

   Priorities of Data Scientist

OR

3. Discuss key elements of effective documentation and code repository aspects of data science.

UNIT-II

4. Explain the information conveyed by a Box Plot with a suitable example. Give some functions available in Python to customize a few aspects of Box Plot.

OR

5. Highlight the differences between univariate and multivariate exploratory data analysis (EDA) techniques with suitable examples.

UNIT-III

6. Describe symmetric and asymmetric distributions, and provide visual representations to illustrate each concept clearly.

OR

7. Elaborate on the concepts of one-way and two-way ANOVA in statistical analysis with suitable examples.

UNIT-IV

8. In exploring the functionalities of Weka Explorer, how does it handle classification and clustering tasks? Could you provide insights into how these tasks are executed within the software?

OR

9. What distinguishes the features of PyBrain, and how does it compare and contrast with Pylearn2 in terms of functionality and capabilities?

UNIT-V

10. Imagine you're embarking on a data science project. What fundamental principles should guide your approach and decision-making throughout the project's lifecycle?

OR

11. Write short notes on:

    a) Seven principles of GDPR

    b) Five features of ethics in data science