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

VI 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) Define the term data science with a neat diagram.

   b) Give the ways of identifying outliers in a dataset using the interquartile range (IQR) method.

   c) Present the main differentiation between skewness and kurtosis.

   d) What are the data formats supported by Weka?

   e) Define the following in ethics in data science
   - Transparency
   - Accountability

Section-B

Answer the following: **(5×5=25)**

UNIT-I

2. Imagine you're exploring the landscape of data science, particularly focusing on large datasets. What are the defining features or traits that distinguish big data from smaller-scale datasets?

OR

3. Investigate the distinction between primary and secondary data sources within the realm of data science.

UNIT-II

4. Write short notes on the following customizing aspects of graphical plots in Python using matplotlib

- Adding the title of the graph
- Adding labels to x and y axes
- Display the legend at the bottom left corner
- Adding ticks to the x-axis
- Making subplots of size 2 x 2

OR

5. Distinguish between univariate and multivariate exploratory data analysis (EDA) as well as graphical and non-graphical approaches to EDA.

UNIT-III

6. Differentiate between covariance and correlation, including the range of correlation values and their interpretation.

OR

7. Illustrate the utility of pivot tables in data analysis, providing a relevant example to demonstrate its functionality.

UNIT-IV

8. Detail the primary components found within five sub-packages of SciPy, outlining their respective functionalities.

OR

9. Elaborate on the core functionalities and standout attributes of Scilab, highlighting its main characteristics.

UNIT-V

10. Outline the core aspects of ethical considerations within the realm of data science, focusing on the five primary features that define ethical practice.

OR

11. Explain the following

- Data Science in Travelling
- Data Science in Banking and Finance