[Apr-24]

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

VI Semester

INTRODUCTION TO DATA SCIENCE
(Offered Through MOOC-Coursera)

(Effective from the admitted batch 2021-22)

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

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

Section-A

1. **Answer all questions:** **(5×1=5)**

   a) Name two common methods for Working with Varied Data Sources and Types.

   b) When is it appropriate to use a line plot for data visualization?

   c) How does Python handle data pre-processing tasks such as scaling and transformation?

   d) What is the key difference between linear regression and multiple linear regression?

   e) What is the primary purpose of using Jupyter Notebooks in data science projects?

Section-B

Answer any five questions: **(5×5=25)**

2. Compare and contrast the advantages and disadvantages of storing data in traditional on-premise servers versus using Cloud-based storage solutions for Data Science projects. Provide real-world examples to support your arguments.

3. List three common programming languages used in Data Science and briefly explain their respective advantages.

4. Explore the concept of faceting in Shiny applications and how it can be used to create multi-panel visualizations.

5. Discuss advanced customization techniques in ggplot2, such as custom themes, annotations, and scales.

6. Using Python, how would you detect outliers within a given dataset? Provide a step-by-step explanation of the process and demonstrate the implementation of outlier detection techniques using Python libraries such as NumPy and pandas.

7. Compare and contrast in-sample evaluation measures such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

8. Discuss the significance of libraries such as NumPy, Pandas, Matplotlib, and Seaborn in Python for tasks such as numerical computations, data manipulation, and visualization

9. Describe the process of implementing grid search for optimizing hyper parameters in regression methods.