

[Apr-24]

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

VI SEMESTER

INTRODUCTION TO DATA SCIENCE
(Offered Through MOOC-Coursera)

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 essential skills for Data Scientists.
- b) Define faceting in the context of data visualization.
- c) How are missing values typically represented in Python data structures?
- d) Define underfitting in the context of model development.
- e) In R, what function from the ggplot2 package is commonly used to create scatter plots.

Section-B

Answer any five questions: (5×5=25)

- 2. Analyze the ethical considerations surrounding the collection and utilization of data in Data Science projects. How can data scientists ensure responsible and ethical practices in their work?
- 3. Analyze the role of Big Data processing tools such as Hadoop, HDFS, Hive, and Spark in handling large-scale datasets. Discuss their strengths, weaknesses, and scenarios where each tool is most suitable.

4. Describe the key differences between bar charts, histograms, and pie charts. Provide examples of scenarios where each of these visualization techniques would be most appropriate, and justify your choices based on the nature of the data being visualized.
5. Explore how Shiny applications can seamlessly integrate data visualization with statistical analysis.
6. Give examples of Python code demonstrating how to format dates and strings, and how to normalize numerical data using Min-Max scaling.
7. Explain how techniques like feature selection and feature engineering can improve model performance.
8. Explain the significance of open-source tools in the field of data science. Compare and contrast at least two widely used open-source tools for data science, highlighting their key features, advantages, and limitations.
9. Describe the process of implementing grid search for optimizing hyper parameters in regression methods.