AR 24 Reg. No | | | | | | | | | | |

GANDHI INSTITUTE OF ENGINEERING AND TECHNOLOGY UNIVERSITY, ODISHA, GUNUPUR

(GIET UNIVERSITY)

Ph.D. (First Semester) Examinations, December – 2024

23SPPECS1013 - Data Science

(CSE)

Time: 3 hrs Maximum: 70 Marks

The figures in the right hand margin indicate marks.

Answer *ANY FIVE* Questions. **(14 x 5 = 70 Marks)** **Marks**

1.a. Provide definitions for key terms commonly used in data science. In what ways does standardized terminology improve collaboration and communication in data-driven projects? 7

b. Describe the process of data exploration and preparation for analysis. Why is it crucial to perform data cleaning and preprocessing before using it in data science projects? 7

2.a Discuss the various types of data encountered in data science, highlighting the characteristics and sources of each type, such as structured, unstructured, and semi-structured data. 7

b. Explain the tools used in data science. How do these tools help in collecting, analyzing, and visualizing data? Give examples of popular tools and their practical applications. 7

3.a. Examine the methods and technologies used for data storage and management. What benefits and challenges are associated with approaches like databases and data lakes? 7

b. Describe the process of data collection, emphasizing its importance in decision-making. How do APIs (Application Programming Interfaces) facilitate data collection, and what advantages do they offer? 7

4.a Define data governance and discuss its significance in data management. How do strong governance policies help maintain data integrity, security, and regulatory compliance? 7

b. Explain the concept of variance in statistics. How does variance help in understanding the spread or dispersion of data? 7

5.a. Clarify the Central Limit Theorem (CLT) and its importance in statistical analysis. How does it facilitate drawing inferences from a sample to the broader population? 7

b. What is data visualization, and why is it important? Explain how visualizing data helps in understanding complex information. Provide examples of different types of data visualizations and explain when each type is most effective. 7

6.a. Examine the principles of data encoding in visualizations. What role do retinal variables play in shaping how visual information is perceived? 7

b.	Explain the role of Bokeh in data visualization using Python. What are the features and capabilities of Bokeh that make it suitable for creating interactive visualizations?	7
7.a	Compare the features of Bokeh with other Python visualization libraries. How does Bokeh enable customization and interactive visualizations?	7
b.	What are some recent trends in data collection techniques? Explain how methods like IoT (Internet of Things), sensors, and social media data mining are shaping the landscape of data acquisition.	7
8.a.	Evaluate the role of ethics and regulations in data science. Why is adherence to ethical standards and legal requirements essential for data scientists?	7
b.	Discuss various visualization techniques that are gaining prominence in data science. What are the benefits of techniques such as interactive dashboards, 3D visualizations, and geospatial mapping in conveying complex data insights?	7

---End of Paper---