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

**GANDHI INSTITUTE OF ENGINEERING AND TECHNOLOGY UNIVERSITY, ODISHA, GUNUPUR**

# (GIET UNIVERSITY)

Ph.D. (First Semester-Winter) Examinations, June – 2025

## 23WPPECS1013 – 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. | What is the significance of standardized terminology in data science, and how does a common vocabulary contribute to better collaboration and understanding among interdisciplinary teams? | 7 |
| b. | Outline the major tools and platforms used in data science workflows. In what ways do these tools streamline the processes of data collection, cleansing, transformation, and visualization? | 7 |
| 2. a | Classify the different data types commonly encountered in data science. How do structured, semi-structured, and unstructured data differ in their format, source, and usage? | 7 |
| b. | Why are statistical measures such as mean, median, and mode considered foundational in data analysis? Illustrate with examples how these metrics are applied in real-world decision-making. | 7 |
| 3.a. | Detail the typical steps involved in the data preprocessing phase of a data science project. Why is it crucial to explore, clean, and transform data before analysis? | 7 |
| b. | Compare and contrast various data storage and management solutions used in modern data science. What benefits and limitations are associated with approaches like relational databases, NoSQL systems, and data lakes? | 7 |
| 4.a. | How does incorporating data from multiple sources improve the quality and depth of analysis? Provide examples where merging heterogeneous datasets yielded significant business or research outcomes. | 7 |
| b. | Define data governance and discuss its role in organizational data management. How do governance policies ensure data quality, compliance, and accountability? | 7 |
| 5.a. | What is the Central Limit Theorem, and why is it a critical concept in inferential statistics? How does it allow data scientists to make predictions about a population from sample data? | 7 |
| b. | Analyze the characteristics of various data distributions commonly seen in statistical analysis. How do these distributions affect the interpretation and outcome of analytical methods? | 7 |
| 6.a. | What are the most commonly used data visualization methods in data science, and how should one choose the appropriate technique depending on the nature of the data and insights required? | 7 |

| | | |
|---|---|---|
| b. | Explain how data encoding principles and visual variables (such as size, shape, color, and position) influence human perception in data visualizations. Why is effective encoding essential for clarity and comprehension? | 7 |
| 7.a. | Why is it essential to utilize data visualization technologies in analytics? Discuss how modern visualization tools assist in uncovering patterns and enhancing data-driven decision-making. | 7 |
| b. | What makes Bokeh a powerful Python library for data visualization? Compare its capabilities in terms of interactivity, customization, and usability with other libraries like Matplotlib and Plotly. | 7 |
| 8.a. | Discuss how software development practices such as Agile methodology, DevOps, and cloud computing accelerate the creation and deployment of data science solutions. What advantages do these practices offer in project execution? | 7 |
| b. | In the context of data science, why is it vital to address ethical concerns and regulatory requirements? Discuss the implications of ignoring data ethics and the importance of laws like GDPR in protecting user privacy. | 7 |

---End of Paper---