

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY-GURUJADA VIZIANAGARAM
II B. Tech II Semester Supplementary Examinations NOV-2025
INTRODUCTION TO DATA SCIENCE
(DEPARTMENT OF CSE (AI&DS), AI&DS)

Time: 3 hours

Max. Marks: 70

The Question paper consists of Part A & Part B.

Part A is compulsory, Answer all questions. Part B Answers any one question from each unit.

- | | | |
|---------------------------------|--|-----------|
| 1 | PART-A | (20Marks) |
| a) | Define Data Science and list its major applications. | [2] |
| b) | What are the key benefits of implementing a Data Science process? | [2] |
| c) | What is semi-supervised learning? Give one example. | [2] |
| d) | List any two common problems encountered in handling large data. | [2] |
| e) | Define CAP theorem and explain its significance in distributed systems. | [2] |
| f) | Differentiate between ACID and BASE principles. | [2] |
| g) | What is Neo4j? Mention its use in data science. | [2] |
| h) | Define text mining. List two Python libraries used for text analytics. | [2] |
| i) | What are the advantages of data visualization in data science? | [2] |
| j) | Name any two popular dashboard development tools. | [2] |
| PART-B | | |
| (50Marks) | | |
| Question from Unit - I | | |
| 2 | a) Explain the Data Science process and its different stages. | [5] |
| | b) Describe how exploratory data analysis (EDA) contributes to model building. | [5] |
| (OR) | | |
| 3 | a) Discuss the importance of defining project goals and preparing a project charter in the Data Science workflow. | [5] |
| | b) Explain data retrieval, cleansing, and transformation steps in detail. | [5] |
| Question from Unit - II | | |
| 4 | a) Explain the role of Machine Learning in Data Science and describe different types of ML. | [5] |
| | b) Discuss the process of feature engineering, model selection, and prediction using Python tools like scikit-learn. | [5] |
| (OR) | | |
| 5 | a) Discuss techniques for handling large datasets efficiently. | [5] |
| | b) Write short notes on the use of scikit-learn in building predictive models. | [5] |
| Question from Unit - III | | |
| 6 | a) Explain the need for NoSQL databases in the Big Data ecosystem. | [5] |
| | b) Discuss the CAP theorem and the BASE principle with examples. | [5] |
| (OR) | | |
| 7 | a) Compare and contrast relational databases and NoSQL databases. | [5] |
| | b) Discuss the architecture and functioning of the Hadoop framework | [5] |
| Question from Unit - IV | | |
| 8 | a) Explain the use of graph databases in data science and discuss the Cypher query language with examples. | [5] |
| | b) Describe the applications of text mining and analytics using Python libraries like NLTK and SQLite. | [5] |
| (OR) | | |
| 9 | a) Discuss the role of Neo4j in handling graph-based data. | [5] |
| | b) Explain how Python libraries like NLTK and SQLite are used for text data analysis. | [5] |

Question from **Unit - V**

- 10 a) Define data visualization and explain different types of visualization techniques used in data analysis. [5]
b) Discuss the advantages of interactive dashboards in real-time data monitoring. [5]
- (OR)
- 11 a) What are the major components of a data dashboard? [5]
b) Explain the process of developing an interactive visualization using **D3.js** and **Cross filter**. [5]
