(NO. OF PAGES: 2)

(FOR THE CANDIDATES ADMITTED DURING THE ACADEMIC YEAR 2024-2025 ONLY)

24UDA4A1

REG.NO

N.G.M.COLLEGE (AUTONOMOUS) : POLLACHI

END-OF-SEMESTER EXAMINATIONS : MAY- 2026

SEMESTER- IV | TIME : 3 HOURS

BSc Cs With DA | MAXIMUM MARKS: 75

PART-III

24UDA4A1 – INTRODUCTION TO DATA SCIENCE

SECTION – A **(10 X 1 = 10 MARKS)**

ANSWER THE FOLLOWING QUESTIONS.

1. Which of the following is a key benefit of Data Science?

   A) Data storage only B) Extracting insights from data

   C) Reducing hardware cost D) Increasing internet speed

2. EDA stands for _____.

   A) Exploratory Data Analysis b) External Data Access

   c) Error Data Analysis d) Extended Data Application

3. Learning using labelled data is called _____.

   A) Supervised B) Unsupervised C) Semi-supervised D) Reinforcement

4. Which is a faster alternative to MapReduce?

   A) Spark B) SQL C) Python D) Java

5. Predicting diseases using data mainly uses _____.

   A) Machine Learning B) Networking C) Animation D) Graphics

ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES

6. What is Data Science.

7. What is data transformation in the data science process?

8. What is meant by semi-supervised learning**?**

9. Define Hadooop.

10. How is Data Science used in Disease Prediction?

(CONT…2)

Ethical paper

SECTION – B (5 X 5 = 25 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.

11. a) Explain the benefits and uses of Data Science in modern organizations.

(OR)

b) Describe the facets of data with suitable examples.

12. a) Explain the research goals and their importance in the Data Science process.

(OR)

b) Describe the role of Exploratory Data Analysis in understanding datasets..

13. a) Explain the modeling process in machine learning.

(OR)

b) Describe the characteristics of supervised learning algorithms.

14. a) Explain the architecture of the Hadoop framework**.**

(OR)

b) Describe the CAP theorem and its significance in distributed databases**..**

15. a) Explain the steps involved in setting research goals for disease prediction**.**

(OR)

b) Describe the process of data preparation in a disease prediction case study.

SECTION – C (5 X 8 = 40 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.

16. a) Discuss the Data Science process and its major stages with a diagram.

(OR)

b) Analyze the role of Data Science in the Big Data ecosystem.

17. a) Explain the complete Data Science workflow from data retrieval to model building**.**

(OR)

b) Analyze the importance of data transformation techniques in data analysis**.**

18. a) Compare supervised and unsupervised learning algorithms with suitable examples.

(OR)

b) Explain the different types of machine learning algorithms and their applications**.**

19. a) Explain how Apache Spark improves performance compared to MapReduce.

(OR)

b) Discuss the types of NoSQL databases and their applications**.**

20. a) Explain the process of disease profiling using data analytics**.**

(OR)

b) Discuss how data visualization and presentation help in communicating disease prediction results**.**

****

Ethical paper