

KAMARAJ COLLEGE (Autonomous)

Accredited with A+ Grade by NAAC

(Affiliated to Manonmaniam Sundaranar University, Tirunelveli)

(3 Pages)

Reg. No:

Question Code: 26E01107

Course Code: 24UECS31

UG Degree - End Semester Examinations, April 2026

Third Semester

B.Sc., COMPUTER SCIENCE

Introduction to Data Science

(For those who joined in July 2024 onwards)

Time: 3Hours

Maximum: 75 Marks

PART - A (10 × 1 = 10 Marks)

Answer ALL Questions

Choose the correct answer:

- CO:1 1. One major benefit of data science is: _____.
- K:1 (a) Reducing storage (b) Increasing uncertainty
(c) Better predictions (d) Decreasing data
- CO:1 2. In distributed file systems, files are automatically: _____.
- K:1 (a) Replicated (b) Encrypted
(c) Renamed (d) Compressed
- CO:2 3. Internal and external data are types of: _____.
- K:1 (a) Data preparation (b) Data transformation
(c) Data aggregation (d) Data retrieval
- CO:2 4. Building a model is described as: _____.
- K:1 (a) A one-time process (b) An iterative process
(c) A linear process (d) A random process
- CO:3 5. The modeling phase consists of how many main steps?
- K:1 (a) 3 (b) 4
(c) 5 (d) 6
- CO:3 6. Which learning technique uses labeled data to find patterns?
- K:1 (a) Supervised learning (b) Unsupervised learning
(c) Reinforcement learning (d) Semi-supervised learning

- CO:4 7. Which of the following is NOT a component of Hadoop?
K:1 (a) HDFS (b) MapReduce
(c) YARN (d) MySQL
- CO:4 8. The “all or nothing” principle refers to: _____.
K:1 (a) Atomicity (b) Consistency
(c) Isolation (d) Durability
- CO:5 9. Combining tokens to create n-grams is also known as: _____.
K:1 (a) Tokenization (b) Stemming
(c) Shingling (d) Indexing
- CO:5 10. The default tokenizer in Elasticsearch is called: _____.
K:1 (a) Simple tokenizer (b) Standard tokenizer
(c) Bigram tokenizer (d) Custom tokenizer

PART - B (5 X 5 = 25 Marks)

Answer ALL Questions choosing either (a) or (b).

Answer should not exceed 250 words.

- CO:1 11. (a) Illustrate the benefits and uses of data science with suitable
K:3 examples.

(OR)

- (b) Demonstrate the steps involved in the data science process.

- CO:2 12. (a) Analyze the importance of defining research goals in a data
K:4 science project.

(OR)

- (b) Examine the concept of exploratory data analysis with a suitable example.

- CO:3 13. (a) Implement a supervised learning model using labeled data.
K:3

(OR)

- (b) Demonstrate semi-supervised learning with a simple example.

- CO:4 14. (a) Investigate how Spark is faster than MapReduce.
K:4

(OR)

- (b) Differentiate between ACID and BASE principles with suitable analysis.

CO:5 15. (a) Analyze how setting research goals influences disease prediction accuracy.
K:4

(OR)

(b) Examine the need for presentation and automation in disease prediction systems.

PART - C (5 X 8 = 40 Marks)

Answer ALL Questions choosing either (a) or (b).

Answer should not exceed 500 words.

CO:1 16. (a) Assess the various facets of data.

K:5

(OR)

(b) Evaluate the components of the big data ecosystem.

CO:2 17. (a) Apply the steps involved in retrieving data for analysis.

K:3

(OR)

(b) Illustrate the main steps involved in building a model.

CO:3 18. (a) Apply the modeling process steps to build a predictive model.

K:3

(OR)

(b) Implement an unsupervised learning algorithm with an example.

CO:4 19. (a) Analyze the impact of the CAP theorem on distributed systems.
K:4

(OR)

(b) Categorize different types of NoSQL databases.

CO:5 20. (a) Design an effective data retrieval strategy for disease prediction systems.
K:6

(OR)

(b) Develop a data exploration framework to support reliable disease analysis.