



CBCS SCHEME

21CS754

Seventh Semester B.E./B.Tech. Degree Examination, June/July 2026 Introduction to Data Science

Time: 3 hrs.

Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

- 1 a. What is data science? Explain benefits and use of data science. (08 Marks)
b. Explain main categories of data. (12 Marks)

OR

- 2 a. Discuss process of data science. (10 Marks)
b. Explain categories of database. (06 Marks)
c. Write a note on distributed file systems. (04 Marks)

Module-2

- 3 a. Explain process of defining research goal and creating project charter. (08 Marks)
b. Discuss the process of collecting internal and external data. (07 Marks)
c. Explain:
i) Model fit
ii) Predictor variable have a co-efficient
iii) Predictor Significance. (05 Marks)

OR

- 4 a. Explain the following with example:
i) Data entry errors
ii) Redundant whitespaces
iii) Outliers
iv) Missing values (10 Marks)
b. Explain different ways to combine data from different data sources. (10 Marks)

Module-3

- 5 a. Explain python packages to work with data in memory and optimizing operation. (10 Marks)
b. List applications of machine learning in data science. (05 Marks)
c. Explain confusion matrices (05 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and/or equations written eg, 42+8 = 50, will be treated as malpractice.

OR

6 a. Explain:

- i) Training a model
- ii) Validating a model

(10 Marks)

b. With an example explain:

- i) Supervised learning
- ii) Un supervised learning

(10 Marks)

Module-4

7 a. Discuss data visualization options for delivering a dash board to end users.

(10 Marks)

b. Write a Java Script for create table functions

(10 Marks)

OR

8 a. List and explain the reason why you'd create your own custom report instead of available tools.

(09 Marks)

b. Explain reduce count () and reduce sum ()

(11 Marks)

Module-5

9 a. With an example explain how Hadoop achieve parallelism using mapreduce

(10 Marks)

b. Explain Spark frame work.

(10 Marks)

OR

10 a. Apply data science process to assessing risk when loaning money.

(10 Marks)

b. Illustrate the components of SPARK Frame work.

(10 Marks)
