CBCS SCHEME

USN

21CS754

Seventh Semester B.E. Degree Examination, Dec.2025/Jan.2026
Introduction to Data Science

Time: 3 hrs. Max. Marks: 100

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

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.

Module-1

1 a. Illustrate the characteristics of data and the application of big data and data science. **(10 Marks)**

b. Explain the following data types :
i) Audio , Image , Video data ii) Machine Generated data
iii) Graph based data **(10 Marks)**

OR

2 a. Illustrate the following components of big data ecosystem :
i) Distributed file systems ii) NoSQL databases **(10 Marks)**

b. Explain the following data types :
i) Structured data ii) Unstructured data iii) Streaming data. **(10 Marks)**

Module-2

3 a. Develop the research goal and project charter with an example. **(10 Marks)**

b. Summarize the data science process of cleaning the data. **(10 Marks)**

OR

4 a. Develop the linear prediction model on semi random data and analyze the model information output. **(10 Marks)**

b. Summarize the different visualization technique graphs for the data exploration. **(10 Marks)**

Module-3

5 a. Define Machine Learning and explain the application of machine learning in Data Science. **(10 Marks)**

b. Construct the Naïve Bayes classifier model to let a computer recognize images of numbers. **(10 Marks)**

OR

6 a. Explain the different Python tools used in Machine Learning. **(10 Marks)**

b. Construct the unsupervised learning to execute the Principal Component Analysis (PCA) of the Redwine Quality Dataset. **(10 Marks)**

Module-4

7 a. Explain the different ways of data visualization. **(10 Marks)**

b. Case study : Consider in the hospital pharmacy , all medicines should be checked for their light sensitivity. The data set consists of time series data for an entire year of stock for 29 medicines.
Examine the steps involved to create the dashboard for the hospital pharmacy. **(10 Marks)**

OR

8 a. Explain the following data visualization libraries :
i) cross filter. js
ii) d3.js
iii) dc.js **(10 Marks)**

b. Examine cross filter data visualization library to filter the medicine data set. [Consider the Q.7b case study]. **(10 Marks)**

Module-5

9 a. Construct the Hadoop framework to store and process the large data set. **(08 Marks)**

b. Case study : Lending club is an organization that connects people in need of a loan with people who have money to invest. Your boss has money to invest and wants information before throwing a substantial sum on the table. Build a interactive dashboard with the relevant information. Design the data retrieval using Hadoop file system and data preparation using spark. **(12 Marks)**

OR

10 a. Construct the Spark framework to store and process the large data set. **(08 Marks)**

b. Design the data explanation and report building using visualization tool Qlik sense. [for the case study in Q.9b]. **(12 Marks)**

* * * * *