

# CBCS SCHEME



USN

--	--	--	--	--	--	--	--	--	--

21CS644

## Sixth Semester B.E./B.Tech. Degree Examination, June/July 2025 Data Science and Visualization

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 relationship between data science and big data. (10 Marks)  
b. Define model. Explain the how you can build a statistical model. (10 Marks)

OR

- 2 a. Explain the Venn diagram of data science. (10 Marks)  
b. Explain the probability distribution. (05 Marks)  
c. Differentiate population and sample, explain the notion of population and sample in the context of Big Data. (05 Marks)

### Module-2

- 3 a. Explain data science process with neat diagram. (10 Marks)  
b. Consider the following dataset given in table cluster, it using K-means algorithm with the initial value of objects 2 and 5 with the co-ordinate values (4, 6) and (12, 4) as initial seeds.

Objects	x-co-ordinate	y-co-ordinate
1	2	4
2	4	6
3	6	8
4	10	4
5	12	4

(05 Marks)

- c. Explain the linear regression algorithm with R implementation with example. (05 Marks)

OR

- 4 a. Explain EDA with suitable example. (10 Marks)  
b. Explain the K-NN algorithm in detail with example. (10 Marks)

**Module-3**

- 5 a. Assume a student performance during his course of study and predict whether the student will get a Job-offer or not in his final year of the course.

Sl. No.	CGPA	Interactiveness	Practical knowledge	Communication skills	Job-offer
1	$\geq 9$	Yes	V. good	Good	Yes
2	$\geq 8$	No	Good	Medium	Yes
3	$\geq 9$	No	Average	Poor	No
4	$< 8$	No	Average	Good	No
5	$\geq 8$	Yes	Good	Medium	Yes
6	$\geq 9$	Yes	Good	Medium	Yes
7	$< 8$	Yes	Good	Poor	No
8	$\geq 9$	No	V. Good	Poor	Yes
9	$\geq 8$	Yes	Good	Good	Yes
10	$\geq 8$	Yes	Average	Good	Yes

Draw decision tree using CART algorithm with Gini index technique. (10 Marks)

- b. Explain random forest algorithm with example and advantages and disadvantages. (10 Marks)

**OR**

- 6 a. Explain singular value decomposition. (10 Marks)  
 b. Define feature extracting, how to do extracting of features and explain its applications. (05 Marks)  
 c. Briefly explain the attribute selection measure technique with example. (05 Marks)

**Module-4**

- 7 a. Define data visualization, Explain the importance of it. (10 Marks)  
 b. Define Data Wrangling. Explain with suitable example. (10 Marks)

**OR**

- 8 a. Explain the any three relation plots with implementation of example dataset. (10 Marks)  
 b. Explain any two composition plot and distribution plot with neat diagram. (10 Marks)

**Module-5**

- 9 a. How to plot a graph using R—implementation. Explain with proper example. (10 Marks)  
 b. Define scatter plots. Explain types of scatter plots in detail. (05 Marks)  
 c. What is Histogram? How we use Histograms in data science? (05 Marks)

**OR**

- 10 a. Write image operations and explain with example. (05 Marks)  
 b. Explain the following:  
   i) Labels  
   ii) Titles  
   iii) Annotations  
   iv) Legends (05 Marks)  
 c. How to create, display and closing figures and format strings? Explain with proper implementation. (10 Marks)

\*\*\*\*\*