JHARKHAND Rai University
RANCHI

PRACTICE SET
End Semester Examination, Spring- 2026

Program: BCA
Semester: VI
Subject: Fundamental of Data Science
Subject Code: 3CCDE306

Course Outcome:
On the completion of the Course, the students will be able to:

Course Outcomes	Description
CO1	Describe the data science process, its importance, and the ethical challenges in data-driven decision-making.
CO2	Apply data cleaning, normalization, and profiling techniques for effective data preparation.
CO3	Perform statistical analysis using descriptive statistics, hypothesis testing, and regression methods.
CO4	Implement basic machine learning algorithms and evaluate model performance using appropriate metrics.
CO5	Create meaningful visualizations and present data insights using storytelling techniques.

1. **Section A : 10 Marks Question covering All units (Total No. of questions 30)**
2. **Section B : 20 Marks Question covering All units (Total No. of questions 10)**

UNIT I
Section A (10 marks)

1. What is the difference between Data Science and Data Analytics? List any five applications of Data Science. (CO1) (LOT)
2. Why is Data Science important in today's world? List the industries where Data Science is widely used. (CO1) (LOT)
3. Explain the stages of the Data Science life cycle. (CO1) (LOT)
4. You are given incomplete and noisy data. How would you handle it in the Data Science lifecycle? (CO1) (HOT)
5. Compare traditional decision-making vs data-driven decision-making. Which is more reliable and why? (CO1) (HOT)

6. Suppose a company wants to predict customer churn. Explain how the Data Science lifecycle will be applied step by step. (CO1) (HOT)

Section B (20 marks)

7. Analyze how Data Science contributes to solving real-world problems in healthcare or education. (CO1) (HOT)
8. Design a simple Data Science approach for analyzing student performance in a university. (CO1) (HOT)

UNIT II

Section A (10 marks)

9. What is primary and secondary data? Explain different sources of data collection. (CO2) (LOT)
10. Why is data cleaning important? List common data cleaning techniques. (CO2) (LOT)
11. Explain data pre-processing. What are missing data and outliers? Explain briefly. (CO2) (LOT)
12. Differentiate between normalization and standardization with examples. (CO2) (LOT)
13. How would you handle outliers in a financial dataset? Explain with methods. (CO2) (LOT)
14. Evaluate the importance of data preprocessing before model building. (CO2) (HOT)

Section B (20 marks)

15. A dataset contains missing values and duplicate entries. Design a step-by-step data cleaning process. (CO2) (HOT)
16. A company collects customer data from multiple sources with inconsistent formats. Suggest a cleaning and integration approach. Also design a complete data preprocessing pipeline for a machine learning project. (CO2) (HOT)

UNIT III

Section A (10 marks)

17. Define Data Analysis. Explain the types of data analysis? What is the role of statistics in data analysis? (CO3) (LOT)
18. Explain null hypothesis (H_0) and alternative hypothesis (H_1) (CO3) (LOT)
19. What is the difference between correlation and regression? (CO3) (LOT)
20. Define skewness and kurtosis. Also explain mean, median, and mode with an example. (CO3) (LOT)
21. Calculate the mean, median, and mode of the dataset: 10, 15, 20, 20, 25, 30, 35 (CO3) (LOT)
22. Calculate the mean using the following frequency distribution: (CO3) (LOT)

Class Interval	Frequency
0–10	5
10–20	9

Class Interval	Frequency
20–30	6
30–40	10

Section B (20 marks)

23. Apply hypothesis testing to determine whether a new teaching method improves student performance. (CO3) (HOT)
24. Compare and contrast between simple linear regression and multiple regression in practical applications. (CO3) (HOT)

UNIT IV

Section A (10 marks)

25. Define Machine Learning. What is the difference between Machine Learning and traditional programming? (CO4) (LOT)
26. Evaluate the importance of feature selection in Machine Learning. (CO4) (LOT)
27. "More data does not always mean better models." Justify. (CO4) (HOT)
28. What is overfitting? Also explain underfitting. (CO4) (LOT)
29. Define Machine Learning. Explain its different types with suitable examples. (CO4) (LOT)
30. What are labeled and unlabeled datasets? Explain their differences with suitable examples. Also mention their role in supervised and unsupervised learning. (CO4) (LOT)

Section B (20 marks)

31. Analyze the strengths and weaknesses of different ML algorithms. (CO4) (HOT)
32. A model performs well on training data but poorly on test data. Analyze the problem and suggest solutions. (CO4) (HOT)

UNIT V

Section A (10 marks)

33. Define data visualization. What is the purpose of data visualization? (CO5) (LOT)
34. Define infographics. Explain their importance in data visualization and how they help in effective communication of information. Give suitable examples. (CO5) (LOT)
35. You are given a large dataset. How would you choose the right visualization technique? (CO5) (LOT)
36. Explain the role of visual perception in data visualization. How does it influence the interpretation of data? (CO5) (LOT)
37. Name any four data visualization tools. Explain briefly about Tableau, Power BI, Matplotlib, and Seaborn. (CO5) (LOT)
38. What is data storytelling? Explain its importance in effective communication. (CO5) (LOT)

Section B (20 marks)

39. Design an appropriate visualization for sales data over 5 years and justify your choice. (CO5) (HOT)
40. Explain with the help of an example: (CO5) (HOT)
    a. Bar chart
    b. Line graph
    c. Pie chart
    d. Scatter plot
    e. Heatmap

Summary Sheet

CO Wise

CO	Q. No	Marks
CO1	1,2,3,4,5,6,7,8	100
CO2	9,10,11,12,13,14,15,16	100
CO3	17,18,19,20,21,22,23,24	100
CO4	25,26,27,28,29,30,31,32	100
CO5	33,34,35,36,37,38,39,40	100
Total		**500**

Unit Wise

Unit	Q. No	Marks
Unit 1	1,2,3,4,5,6,7,8	100
Unit 2	9,10,11,12,13,14,15,16	100
Unit 3	17,18,19,20,21,22,23,24	100
Unit 4	25,26,27,28,29,30,31,32	100
Unit 5	33,34,35,36,37,38,39,40	100
Total		**500**

Blooms Taxonomy Level (BTL) Wise

BTL	Q. No	Marks
LOT	1,2,3,9,10,11,12,13,17,18,19,20,21,22,25,26,28,29,30,33, 34,35,36,37,38	250
HOT	4,5,6,7,8,14,15,16,23,24,27,31,32,39,40	250
Total		**500**

Prepared By: Dr. Kumar Amrendra

Disclaimer: -This is a Practice set. The Question in End term examination will differ from the Practice set. This Practice set is meant for practice only.