

Code No.: R22DS515PE

R22

H.T.No.

8

R

**CMR ENGINEERING COLLEGE: : HYDERABAD**  
**UGC AUTONOMOUS**

**III–B.TECH–I–Semester End Examinations (Supply) - June- 2025**

**DATA SCIENCE APPLICATIONS**

**(CSD)**

**[Time: 3 Hours]**

**[Max. Marks: 60]**

**Note:** This question paper contains two parts A and B.

Part A is compulsory which carries 10 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks.

**PART-A**

**(10 Marks)**

1. a) Define data science. [1M]
- b) Give the applications of Data science. [1M]
- c) Define data discretization. [1M]
- d) How does data cleaning differ across structured, unstructured data? [1M]
- e) What is mean? [1M]
- f) Define skewness. [1M]
- g) What is polynomial regression? [1M]
- h) What is the impact of multiple regression on model reliability? [1M]
- i) Define grid search. [1M]
- j) What is ridge regression? [1M]

**PART-B**

**(50 Marks)**

2. Discuss the technological advancements that have contributed to the growth of data science. [10M]
- OR**
3. How can organizations balance data utility and security in data science applications? [10M]
4. Explain the concept of data cleaning and its significance in preparing raw data for analysis. [10M]
- OR**
5. Discuss the role of data integration in combining data from multiple sources for analysis. [10M]
6. How do descriptive statistics differ from inferential statistics? Provide examples. [10M]
- OR**
7. Explain the components of a box plot and their significance in visualizing data. [10M]
8. Discuss the role of regression models in making predictions and informed decisions. [10M]
- OR**
9. How can visualization be used to identify potential overfitting or underfitting in regression models? [10M]
10. Discuss the role of cross-validation in model selection and avoiding overfitting. [10M]
- OR**
11. How can strategies such as regularization and cross-validation reduce generalization error? [10M]

\*\*\*\*\*