

Code No.: DS513PE

R20

H.T.No.

8

R

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

**III-B.TECH-I-Semester End Examinations (Regular) - January- 2024
INTRODUCTION TO DATA SCIENCE
(CSD)**

[Time: 3 Hours]

[Max. Marks: 70]

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

Part A is compulsory which carries 20 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

(20 Marks)

1. a) Define Data Science? [2M]
- b) Describe Data security? [2M]
- c) Discuss Data transformation? [2M]
- d) State Outlier? [2M]
- e) What is Skewness? [2M]
- f) Describe Box plot? [2M]
- g) What is distribution plot? [2M]
- h) What is regression line? [2M]
- i) Explain cross validation? [2M]
- j) What is confusion matrix? [2M]

PART-B

(50 Marks)

2. Explain the evolution of data science methodologies in details? [10M]
- OR**
3. Describe the stages in data science project in detail? [10M]
 4. What are the different Data Collection Methods? Explain in detail. [10M]
- OR**
5. What is Data Pre-processing? Discuss about various steps to pre-process the data? [10M]
 6. Describe skewness and kurtosis in descriptive statistics? Explain difference between positive and negative kurtosis. [10M]
- OR**
7. Explain the following with example [10M]
 - i. One way ANOVA
 - ii. Two way ANOVA
 8. What is simple linear regression? Explain assumptions of simple linear regression with example. [10M]
- OR**
9. What is polynomial regression and pipelines in model development? Explain in detail. [10M]
 10. Explain the concept of overfitting and underfitting with examples? How do you strike a balance between them? [10M]
- OR**
11. What is Ridge regression? Explain the difference between ridge regression and ordinary least squares (OLS). [10M]
